Fig. 1 – Sequence information for C-LytA.

SEQ ID NO:1 – amino acid sequence of C-LytA repeat 1

GWQKNDTGYWYVHSD 15

SEQ ID NO:2 – amino acid sequence of C-LytA repeat 2

GSYPKDKFEKINGTWYYFDSS 21

SEQ ID NO:3 – amino acid sequence of C-LytA repeat 3

GYMLADRWRKHTDGNWYWFDNS 22

SEQ ID NO:4 – amino acid sequence of C-LytA repeat 4

GEMATGWKKIADKWYYFNEE 20

SEQ ID NO:5 – amino acid sequence of C-LytA repeat 5

GAMKTGWVKYKDTWYYLDAKE 21

SEQ ID NO:6 – amino acid sequence of C-LytA repeat 6

GAMVSNAFIQSADGTGWYYLKPD 23

SEQ ID NO:7 – amino acid sequence of C-LytA cholin-binding domain

GWQKNDTGYW YVHSDGSYPK DKFEKINGTW YYFDSSGYML ADRWRKHTDG NWYWFDNSGE 60
MATGWKKIAD KWYYFNEEGA MKTGWVKYKD TWYYLDAKEG AMVSNAFIQS ADGTGWYYLK 120
PDGTLADRPE FTVEPDGLIT VK 142

SEQ ID NO:8 – amino acid sequence of C-LytA domain from truncated repeat 1 to repeat 6 (as part of our constructs shown in figure 2)

YVHSD**GSYPKDKFEKINGTWYYFDS**SGYMLADRWRKHTDGNWYWFDNS**GEMATGWKKIADKWYYFNEE**GAMKT
GWVKYKDTWYYLDAKE**GAMVSNAFIQSADGTGWYYLKPD**

SEQ ID NO:9 – DNA sequence encoding the amino acid sequence of SEQ ID NO:1

ggctggcaga agaatgacac tggctactgg tacgtacatt cagac

SEQ ID NO:10 – DNA sequence encoding the amino acid sequence of SEQ ID NO:2

ggctcttatc caaaagacaa gtttgagaaa atcaatggca cttggtacta ctttgacagt tca

SEQ ID NO:11 – DNA sequence encoding the amino acid sequence of SEQ ID NO:3

ggctatatgc ttgcagaccg ctggaggaag cacacagacg gcaactggta ctggttcgac aactca

SEQ ID NO:12 – DNA sequence encoding the amino acid sequence of SEQ ID NO:4

ggcgaaatgg ctacaggctg gaagaaaatc gctgataagt ggtactattt caacgaagaa

SEQ ID NO:13 – DNA sequence encoding the amino acid sequence of SEQ ID NO:5

Ggtgccatga agacaggctg ggtcaagtac aaggacactt ggtactactt agacgctaaa gaa

SEQ ID NO:14 – DNA sequence encoding the amino acid sequence of SEQ ID NO:6

Ggcgccatgg tatcaaatgc ctttatccag tcagcggacg gaacaggctg gtactacctc
aaaccagac

SEQ ID NO:15 – DNA sequence encoding the amino acid sequence of SEQ ID NO:7

ggctggcaga agaatgacac tggctactgg tacgtacatt cagacggctc ttatccaaaa 60
gacaagtttg agaaaatcaa tggcacttgg tactactttg acagttcagg ctatatgctt 120
gcagaccgct ggaggaagca cacagacggc aactggtact ggttcgacaa ctcaggcgaa 180
atggctacag gctggaagaa aatcgctgat aagtggtact atttcaacga agaaggtgcc 240
atgaagacag gctgggtcaa gtacaaggac acttggtact acttagacgc taaagaaggc 300
gccatggtat caaatgcctt tatccagtca gcggacggaa caggctggta ctacctcaaa 360
ccagacggaa cactggcaga caggccagaa ttcacagtag agccagatgg cttgattaca 420
gtaaaataa 429

SEQ ID NO:16 – DNA sequence encoding the amino acid sequence of SEQ ID NO:8

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
GTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTA
TCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGAC

FIG. 2. CPC and native Constructs

Construct 1 – coding sequence of CPC-P501$_{51\text{-}553}$ (see plasmid of figure 7 -Y1796)

Protein sequence (SEQ ID NO:27)

R1 R2 R3 R4

MAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATG

R5 P2 R6

WKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQS
ADGTGWYYLKPDGTLADRPEKFMYMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSL
GILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSV
YAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEG
LSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE
GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPV
AAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSF
LPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAF
LLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSAGGHHHHHH

R1 (plain): aa5-9 (fragment) **R4 (bold)**: aa53-72 P2 (underline): 97-110
R2 (bold): aa10-30 R5 (plain): aa73-93
R3 (plain): aa31-52 **R6a (bold)**: aa94-95 **R6b (bold)**: 113-133

Nucleotide sequence (SEQ ID NO:28)

ATGgcggccgctTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGT
ACTACTTTGACAGTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTT
CGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGT
GCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaat
acatcaaggctaactctaagttcattggtatcactgaaggcgtcATGGTATCAAATGCCTTTATCCAGTCAGC
GGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAaagttcatgtaCatg
GTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTG
GACGCTATGGCCGCCGCCGGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCC
AAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCCTGGAGCTGGCACTGCTCATCCTGGGC
GTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGG
ACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCCTTCATGATCAGTCTTGGGGGCTGCCTGGGCTACCT
CCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCTCTTTGGC
CTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCA
CCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTT

CCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCCGCATGCCCCGCACCCTGCGCCGGCTC
TTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCTGTTTTACACGGATTTCGTGGGCGAGG
GGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGAT
GGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGGTCATGGACCGGCTGGTGCAG
CGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGT
CCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCC
CTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGGAGGT
GCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGAC
ACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCGCGCTCTGCGGGGCCTCTGCCTGTGAtGTCTC
CGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGTGGTTCCGGGCCGGGGCATCTGCCTGGACCTCGCC
ATCCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCC
AGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGGGTCTGGTCGCCATTTACTTTGCTACACAGGTAGT
ATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtggacaccatcaccatcaccattaa

Construct 2 – Coding sequence of P501$_{55-553}$ HIS (control) (yeast strain SC333)

Protein sequence (SEQ ID NO:29)

MVLGIGPVLG LVCVPLLGSA SDHWRGRYGR RRPFIWALSL GILLSLFLIP RAGWLAGLLC 60
PDPRPLELAL LILGVGLLDF CGQVCFTPLE ALLSDLFRDP DHCRQAYSVY AFMISLGGCL 120
GYLLPAIDWD TSALAPYLGT QEECLFGLLT LIFLTCVAAT LLVAEEAALG PTEPAEGLSA 180
PSLSPHCCPC RARLAFRNLG ALLPRLHQLC CRMPRTLRRL FVAELCSWMA LMTFTLFYTD 240
FVGEGLYQGV PRAEPGTEAR RHYDEGVRMG SLGLFLQCAI SLVFSLVMDR LVQRFGTRAV 300
YLASVAAFPV AAGATCLSHS VAVVTASAAL TGFTFSALQI LPYTLASLYH REKQVFLPKY 360
RGDTGGASSE DSLMTSFLPG PKPGAPFPNG HVGAGGSGLL PPPPALCGAS ACDVSVRVVV 420
GEPTEARVVP GRGICLDLAI LDSAFLLSQV APSLFMGSIV QLSQSVTAYM VSAAGLGLVA 480
IYFATQVVFD KSDLAKYSAG GHHHHHH 507

Nucleotide sequence (SEQ ID NO:30)

atgGTGCTGG GCATTGGTCC AGTGCTGGGC CTGGTCTGTG TCCCGCTCCT AGGCTCAGCC 60
AGTGACCACT GGCGTGGACG CTATGGCCGC CGCCGGCCCT TCATCTGGGC ACTGTCCTTG 120
GGCATCCTGC TGAGCCTCTT TCTCATCCCA AGGGCCGGCT GGCTAGCAGG GCTGCTGTGC 180
CCGGATCCCA GGCCCCTGGA GCTGGCACTG CTCATCCTGG GCGTGGGGCT GCTGGACTTC 240
TGTGGCCAGG TGTGCTTCAC TCCACTGGAG GCCCTGCTCT CTGACCTCTT CCGGGACCCG 300
GACCACTGTC GCCAGGCCTA CTCTGTCTAT GCCTTCATGA TCAGTCTTGG GGGCTGCCTG 360
GGCTACCTCC TGCCTGCCAT TGACTGGGAC ACCAGTGCCC TGGCCCCCTA CCTGGGCACC 420
CAGGAGGAGT GCCTCTTTGG CCTGCTCACC CTCATCTTCC TCACCTGCGT AGCAGCCACA 480
CTGCTGGTGG CTGAGGAGGC AGCGCTGGGC CCCACCGAGC CAGCAGAAGG GCTGTCGGCC 540
CCCTCCTTGT CGCCCCACTG CTGTCCATGC CGGGCCCGCT TGGCTTTCCG GAACCTGGGC 600

Rec'd PCT/PTO 08 DEC 2004 10/517420


```
GCCCTGCTTC CCCGGCTGCA CCAGCTGTGC TGCCGCATGC CCCGCACCCT GCGCCGGCTC   660
TTCGTGGCTG AGCTGTGCAG CTGGATGGCA CTCATGACCT TCACGCTGTT TTACACGGAT   720
TTCGTGGGCG AGGGGCTGTA CCAGGGCGTG CCCAGAGCTG AGCCGGGCAC CGAGGCCCGG   780
AGACACTATG ATGAAGGCGT TCGGATGGGC AGCCTGGGGC TGTTCCTGCA GTGCGCCATC   840
TCCCTGGTCT TCTCTCTGGT CATGGACCGG CTGGTGCAGC GATTCGGCAC TCGAGCAGTC   900
TATTTGGCCA GTGTGGCAGC TTTCCCTGTG GCTGCCGGTG CCACATGCCT GTCCCACAGT   960
GTGGCCGTGG TGACAGCTTC AGCCGCCCTC ACCGGGTTCA CCTTCTCAGC CCTGCAGATC  1020
CTGCCCTACA CACTGGCCTC CCTCTACCAC CGGGAGAAGC AGGTGTTCCT GCCCAAATAC  1080
CGAGGGGACA CTGGAGGTGC TAGCAGTGAG GACAGCCTGA TGACCAGCTT CCTGCCAGGC  1140
CCTAAGCCTG GAGCTCCCTT CCCTAATGGA CACGTGGGTG CTGGAGGCAG TGGCCTGCTC  1200
CCACCTCCAC CCGCGCTCTG CGGGGCCTCT GCCTGTGAtG TCTCCGTACG TGTGGTGGTG  1260
GGTGAGCCCA CCGAGGCCAG GGTGGTTCCG GGCCGGGGCA TCTGCCTGGA CCTCGCCATC  1320
CTGGATAGTG CCTTCCTGCT GTCCCAGGTG GCCCCATCCC TGTTTATGGG CTCCATTGTC  1380
CAGCTCAGCC AGTCTGTCAC TGCCTATATG GTGTCTGCCG CAGGCCTGGG TCTGGTCGCC  1440
ATTTACTTTG CTACACAGGT AGTATTTGAC AAGAGCGACT TGGCCAAATA CTCAGCGggt  1500
ggacaccatc accatcacca ttaa  1524
```

Construct 3 - Coding sequence of natssP501$_{1\text{-}34}$ P501$_{51\text{-}553}$ HIS (yeast strain Y1800)

Protein sequence (SEQ ID NO:31)

R1 R2

MAAVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAAYVHSDGSYPKDKFEKINGTW

R3 R4 R5

YYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYYFNEEGAMKTGWVK

P2 R6

YKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQSADGTGWYYLKPDGTLADRPEKFMY
MVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLEL
ALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAP
YLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPR
LHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMG
SLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSA
LQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPA
LCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVS
AAGLGLVAIYFATQVVFDKSDLAKYSAGGHHHHHH

R1 (plain): aa38-42 (fragment) **R4 (bold)**: aa77-106 P2 (underline): 130-143

R2 (bold): aa43-64 R5 (plain): aa107-126

R3 (plain): aa65-76 **R6a (bold)**: aa127-128 **R6b (bold)**: aa146-166

natss stands for native signal sequence

Nucleotide sequence (SEQ ID NO:32)

ATGgcGGCCGTGCAGAGGCTATGGGTATCGAGACTGCTAAGACACCGCAAAGCTCAGTTGTTGTTGGTTAACT
TGTTGACCTTCGGGCTGGAAGTCTGTTTGGCggccgctTACGTACATTCCGACGGCTCTTATCCAAAAGACAA
GTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAG
CACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATA
AGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTT
AGACGCTAAAGAAGGCGCCatgcaatacatcaaggctaactctaagttcattggtatcactgaaggcgtcATG
GTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAG
ACAGGCCAGAAaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCT
AGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGGCCCTTCATCTGGGCACTGTCCTTGGGC
ATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCC
TGGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGA
GGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCCTTCATGATC
AGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGG
GCACCCAGGAGGAGTGCCTCTTTGGCCTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACACTGCTGGT
GGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCACTGC
TGTCCATGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCC
GCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCT
GTTTTACACGGATTTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGG
AGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCT
CTCTGGTCATGGACCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCC
TGTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTC
ACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGC
CCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAA
GCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCGCGCTC
TGCGGGGCCTCTGCCTGTGAtGTCTCCGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGTGGTTCCGG
GCCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCATCCCTGTT
TATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGGGTCTGGTC
GCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtggacaccatc
accatcaccattaa

Construct 4 - Coding sequence of alphapreCPC-P501$_{51-553}$ HIS (yeast strain Y1802)

Protein sequence (SEQ ID NO:33)

Alpha-pre signal R1 R2 R3
MAARFPSIFTAVLFAASSALAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFD

R4 R5 P2
NSGEMATGWKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFI

R6

$\boxed{\text{QSADGTGWYYLKPD}}$GTLADRPEKFMYMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLF
LIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCL
GYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRAR
LAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEG
VRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQ
ILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASAC
DVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFAT
QVVFDKSDLAKYSAGGHHHHHH

Alpha-pre signal (bold): aa4-22

R1 (plain): aa24-28 (fragment)

R2 (bold): aa29-49

R3 (plain): aa50-71

R4 (bold): aa72-91

R5 (plain): aa92-112

R6a (bold): aa113-114

P2 (underline): 116-129

R6b (bold): aa132-152

Alphapre stands for alpha pre signal sequence

Nucleotide sequence (SEQ ID NO:34)

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
GTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatg<u>caatacatcaaggcta</u>
<u>actctaagttcattggtatcactgaa</u>ggcgtcATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGG
CTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAA

ATGgcGGCCAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCATTAGCggccgctTACG
TACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTC
AGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAA
ATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCT
GGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatg<u>caatacatcaaggctaactc</u>
<u>taagttcattggtatcactgaa</u>ggcgtcATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGG
TACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAgctggtattacttacgttccaccattgttgt
tggaagttggtgttgaagaaaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGT
CCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGGCCCTTCATCTGGGCACTG
TCCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATC
CCAGGCCCCTGGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCAC
TCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCT
TCATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCC
CTACCTGGGCACCCAGGAGGAGTGCCTCTTTGGCCTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACA

CTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTCGC
CCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCT
GTGCTGCCGCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACC
TTCACGCTGTTTTACACGGATTTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCG
AGGCCCGGAGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCT
GGTCTTCTCTCTGGTCATGGACCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCA
GCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCA
CCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGT
GTTCCTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCA
GGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCAC
CCGCGCTCTGCGGGGCCTCTGCCTGTGAtGTCTCCGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGT
GGTTCCGGGCCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCA
TCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGG
GTCTGGTCGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtgg
acaccatcaccatcaccattaa

<u>Construct 5 - Coding sequence of alphaprepro-P501$_{51-553}$ HIS (in plasmid pRIT 15068 and yeast strain Y1790)</u>

Protein sequence (SEQ ID NO:35)

```
MSFLNFTAVL FAASSALAAP VNTTTEDETA QIPAEAVIGY SDLEGDFDVA VLPFSNSTNN  60
GLLFINTTIA SIAAKEEGVS LEKREAEAMV LGIGPVLGLV CVPLLGSASD HWRGRYGRRR  120
PFIWALSLGI LLSLFLIPRA GWLAGLLCPD PRPLELALLI LGVGLLDFCG QVCFTPLEAL  180
LSDLFRDPDH CRQAYSVYAF MISLGGCLGY LLPAIDWDTS ALAPYLGTQE ECLFGLLTLI  240
FLTCVAATLL VAEEAALGPT EPAEGLSAPS LSPHCCPCRA RLAFRNLGAL LPRLHQLCCR  300
MPRTLRRLFV AELCSWMALM TFTLFYTDFV GEGLYQGVPR AEPGTEARRH YDEGVRMGSL  360
GLFLQCAISL VFSLVMDRLV QRFGTRAVYL ASVAAFPVAA GATCLSHSVA VVTASAALTG  420
FTFSALQILP YTLASLYHRE KQVFLPKYRG DTGGASSEDS LMTSFLPGPK PGAPFPNGHV  480
GAGGSGLLPP PPALCGASAC DVSVRVVVGE PTEARVVPGR GICLDLAILD SAFLLSQVAP  540
SLFMGSIVQL SQSVTAYMVS AAGLGLVAIY FATQVVFDKS DLAKYSAGGH HHHHH  595
```

Nucleotide sequence (SEQ ID NO:36)

```
ATGAGTTTCC TCAATTTTAC TGCAGTTTTA TTCGCAGCAT CCTCCGCATT AGCTGCTCCA  60
GTCAACACTA CAACAGAAGA TGAAACGGCA CAAATTCCGG CTGAAGCTGT CATCGGTTAC  120
TCAGATTTAG AAGGGGATTT CGATGTTGCT GTTTTGCCAT TTTCCAACAG CACAAATAAC  180
GGGTTATTGT TTATAAATAC TACTATTGCC AGCATTGCTG CTAAAGAAGA AGGGGTATCT  240
CTCGAGAAAA GAGAGGCTGA AGCCatgGTG CTGGGCATTG GTCCAGTGCT GGGCCTGGTC  300
TGTGTCCCGC TCCTAGGCTC AGCCAGTGAC CACTGGCGTG GACGCTATGG CCGCCGCCGG  360
```

CCCTTCATCT GGGCACTGTC CTTGGGCATC CTGCTGAGCC TCTTTCTCAT CCCAAGGGCC 420
GGCTGGCTAG CAGGGCTGCT GTGCCCGGAT CCCAGGCCCC TGGAGCTGGC ACTGCTCATC 480
CTGGGCGTGG GGCTGCTGGA CTTCTGTGGC CAGGTGTGCT TCACTCCACT GGAGGCCCTG 540
CTCTCTGACC TCTTCCGGGA CCCGGACCAC TGTCGCCAGG CCTACTCTGT CTATGCCTTC 600
ATGATCAGTC TTGGGGGCTG CCTGGGCTAC CTCCTGCCTG CCATTGACTG GGACACCAGT 660
GCCCTGGCCC CCTACCTGGG CACCCAGGAG GAGTGCCTCT TTGGCCTGCT CACCCTCATC 720
TTCCTCACCT GCGTAGCAGC CACACTGCTG GTGGCTGAGG AGGCAGCGCT GGGCCCCACC 780
GAGCCAGCAG AAGGGCTGTC GGCCCCCTCC TTGTCGCCCC ACTGCTGTCC ATGCCGGGCC 840
CGCTTGGCTT TCCGGAACCT GGGCGCCCTG CTTCCCCGGC TGCACCAGCT GTGCTGCCGC 900
ATGCCCCGCA CCCTGCGCCG GCTCTTCGTG GCTGAGCTGT GCAGCTGGAT GGCACTCATG 960
ACCTTCACGC TGTTTTACAC GGATTTCGTG GGCGAGGGGC TGTACCAGGG CGTGCCCAGA 1020
GCTGAGCCGG GCACCGAGGC CCGGAGACAC TATGATGAAG GCGTTCGGAT GGGCAGCCTG 1080
GGGCTGTTCC TGCAGTGCGC CATCTCCCTG GTCTTCTCTC TGGTCATGGA CCGGCTGGTG 1140
CAGCGATTCG GCACTCGAGC AGTCTATTTG GCCAGTGTGG CAGCTTTCCC TGTGGCTGCC 1200
GGTGCCACAT GCCTGTCCCA CAGTGTGGCC GTGGTGACAG CTTCAGCCGC CCTCACCGGG 1260
TTCACCTTCT CAGCCCTGCA GATCCTGCCC TACACACTGG CCTCCCTCTA CCACCGGGAG 1320
AAGCAGGTGT TCCTGCCCAA ATACCGAGGG GACACTGGAG GTGCTAGCAG TGAGGACAGC 1380
CTGATGACCA GCTTCCTGCC AGGCCCTAAG CCTGGAGCTC CCTTCCCTAA TGGACACGTG 1440
GGTGCTGGAG GCAGTGGCCT GCTCCCACCT CCACCCGCGC TCTGCGGGGC CTCTGCCTGT 1500
GAtGTCTCCG TACGTGTGGT GGTGGGTGAG CCCACCGAGG CCAGGGTGGT TCCGGGCCGG 1560
GGCATCTGCC TGGACCTCGC CATCCTGGAT AGTGCCTTCC TGCTGTCCCA GGTGGCCCCA 1620
TCCCTGTTTA TGGGCTCCAT TGTCCAGCTC AGCCAGTCTG TCACTGCCTA TATGGTGTCT 1680
GCCGCAGGCC TGGGTCTGGT CGCCATTTAC TTTGCTACAC AGGTAGTATT TGACAAGAGC 1740
GACTTGGCCA AATACTCAGC Gggtggacac catcaccatc accattaa 1788

FIG. 3. Structure of CPC-p501 His fusion protein expressed in *S. cerevisiae*

⬭	Clyta repeats
☐	P2 peptide
—	P501 sequences

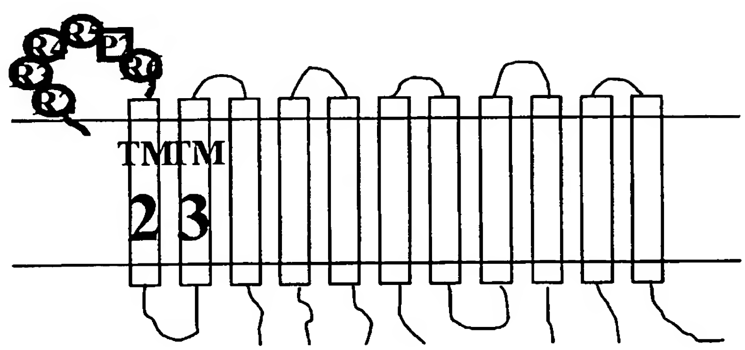


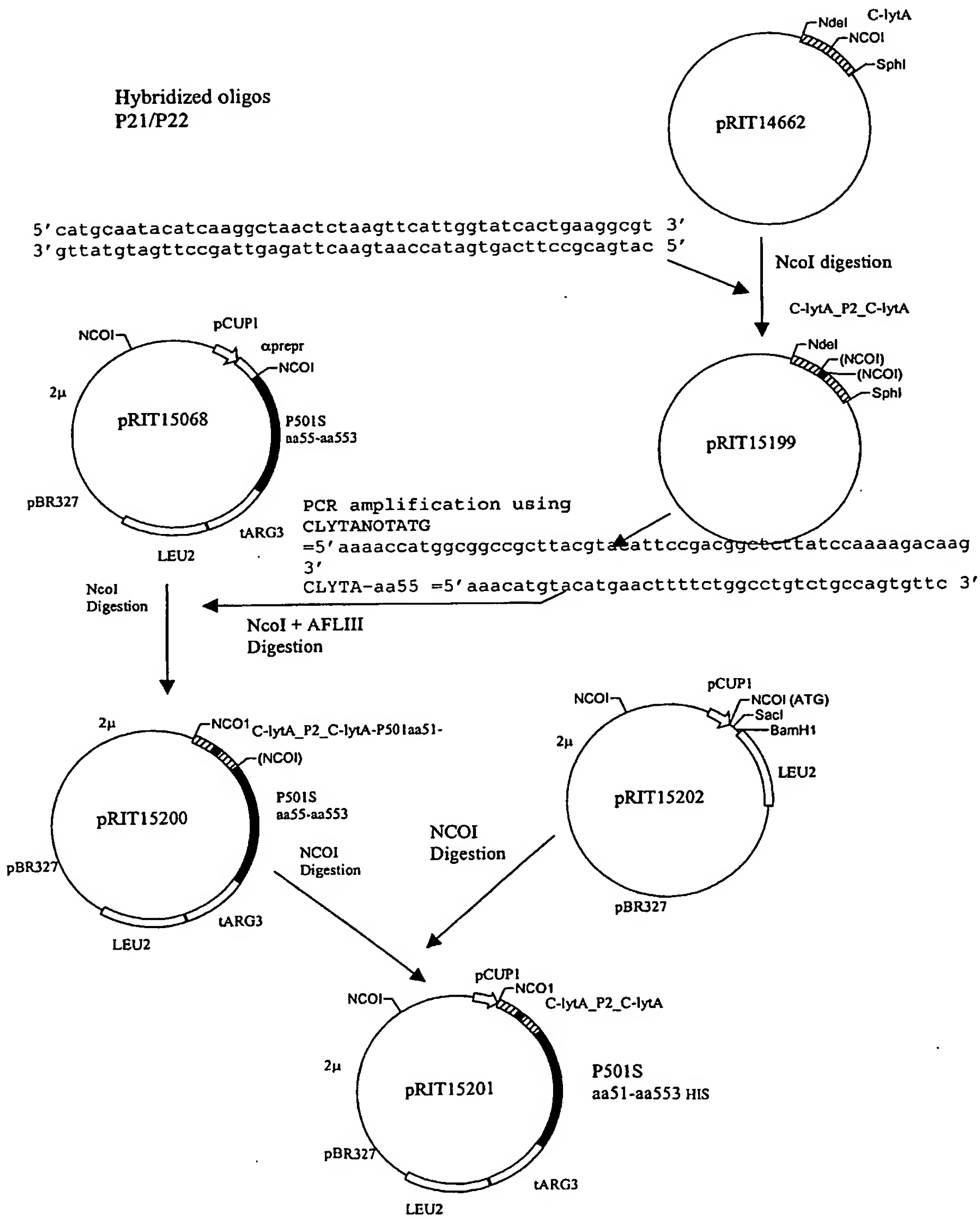
FIG. 4. Primary structure of CPC-P501 His fusion protein (SEQ ID NO.41)

```
MAAAYVHSDG SYPKDKFEKI NGTWYYFDSS GYMLADRWRK HTDGNWYWFD NSGEMATGWK  60
KIADKWYYFN EEGAMKTGWV KYKDTWYYLD AKEGAMQYIK ANSKFIGITE GVMVSNAFIQ  120
SADGTGWYYL KPDGTLADRP EKFMYMVLGI GPVLGLVCVP LLGSASDHWR GRYGRRRPFI  180
WALSLGILLS LFLIPRAGWL AGLLCPDPRP LELALLILGV GLLDFCGQVC FTPLEALLSD  240
LFRDPDHCRQ AYSVYAFMIS LGGCLGYLLP AIDWDTSALA PYLGTQEECL FGLLTLIFLT  300
CVAATLLVAE EAALGPTEPA EGLSAPSLSP HCCPCRARLA FRNLGALLPR LHQLCCRMPR  360
TLRRLFVAEL CSWMALMTFT LFYTDFVGEG LYQGVPRAEP GTEARRHYDE GVRMGSLGLF  420
LQCAISLVFS LVMDRLVQRF GTRAVYLASV AAFPVAAGAT CLSHSVAVVT ASAALTGFTF  480
SALQILPYTL ASLYHREKQV FLPKYRGDTG GASSEDSLMT SFLPGPKPGA PFPNGHVGAG  540
GSGLLPPPPA LCGASACDVS VRVVVGEPTE ARVVPGRGIC LDLAILDSAF LLSQVAPSLF  600
MGSIVQLSQS VTAYMVSAAG LGLVAIYFAT QVVFDKSDLA KYSAGGHHHH HH  652
```

FIG. 5. Nucleotide sequence of CPC P501 His(pRIT15201) (SEQ ID NO.42)

```
ATGGCGGCCG CTTACGTACA TTCCGACGGC TCTTATCCAA AAGACAAGTT TGAGAAAATC   60
AATGGCACTT GGTACTACTT TGACAGTTCA GGCTATATGC TTGCAGACCG CTGGAGGAAG  120
CACACAGACG GCAACTGGTA CTGGTTCGAC AACTCAGGCG AAATGGCTAC AGGCTGGAAG  180
AAAATCGCTG ATAAGTGGTA CTATTTCAAC GAAGAAGGTG CCATGAAGAC AGGCTGGGTC  240
AAGTACAAGG ACACTTGGTA CTACTTAGAC GCTAAAGAAG GCGCCATGCA ATACATCAAG  300
GCTAACTCTA AGTTCATTGG TATCACTGAA GGCGTCATGG TATCAAATGC CTTTATCCAG  360
TCAGCGGACG GAACAGGCTG GTACTACCTC AAACCAGACG GAACACTGGC AGACAGGCCA  420
GAAAAGTTCA TGTACATGGT GCTGGGCATT GGTCCAGTGC TGGGCCTGGT CTGTGTCCCG  480
CTCCTAGGCT CAGCCAGTGA CCACTGGCGT GGACGCTATG GCCGCCGCCG GCCCTTCATC  540
TGGGCACTGT CCTTGGGCAT CCTGCTGAGC CTCTTTCTCA TCCCAAGGGC CGGCTGGCTA  600
GCAGGGCTGC TGTGCCCGGA TCCCAGGCCC CTGGAGCTGG CACTGCTCAT CCTGGGCGTG  660
GGGCTGCTGG ACTTCTGTGG CCAGGTGTGC TTCACTCCAC TGGAGGCCCT GCTCTCTGAC  720
CTCTTCCGGG ACCCGGACCA CTGTCGCCAG GCCTACTCTG TCTATGCCTT CATGATCAGT  780
CTTGGGGGCT GCCTGGGCTA CCTCCTGCCT GCCATTGACT GGGACACCAG TGCCCTGGCC  840
CCCTACCTGG GCACCCAGGA GGAGTGCCTC TTTGGCCTGC TCACCCTCAT CTTCCTCACC  900
TGCGTAGCAG CCACACTGCT GGTGGCTGAG GAGGCAGCGC TGGGCCCCAC CGAGCCAGCA  960
GAAGGGCTGT CGGCCCCCTC CTTGTCGCCC CACTGCTGTC CATGCCGGGC CCGCTTGGCT 1020
TTCCGGAACC TGGGCGCCCT GCTTCCCCGG CTGCACCAGC TGTGCTGCCG CATGCCCCGC 1080
ACCCTGCGCC GGCTCTTCGT GGCTGAGCTG TGCAGCTGGA TGGCACTCAT GACCTTCACG 1140
CTGTTTTACA CGGATTTCGT GGGCGAGGGG CTGTACCAGG GCGTGCCCAG AGCTGAGCCG 1200
GGCACCGAGG CCCGGAGACA CTATGATGAA GGCGTTCGGA TGGGCAGCCT GGGGCTGTTC 1260
CTGCAGTGCG CCATCTCCCT GGTCTTCTCT CTGGTCATGG ACCGGCTGGT GCAGCGATTC 1320
GGCACTCGAG CAGTCTATTT GGCCAGTGTG GCAGCTTTCC CTGTGGCTGC CGGTGCCACA 1380
TGCCTGTCCC ACAGTGTGGC CGTGGTGACA GCTTCAGCCG CCCTCACCGG GTTCACCTTC 1440
TCAGCCCTGC AGATCCTGCC CTACACACTG GCCTCCCTCT ACCACCGGGA GAAGCAGGTG 1500
TTCCTGCCCA AATACCGAGG GGACACTGGA GGTGCTAGCA GTGAGGACAG CCTGATGACC 1560
AGCTTCCTGC CAGGCCCTAA GCCTGGAGCT CCCTTCCCTA ATGGACACGT GGGTGCTGGA 1620
GGCAGTGGCC TGCTCCCACC TCCACCCGCG CTCTGCGGGG CCTCTGCCTG TGATGTCTCC 1680
GTACGTGTGG TGGTGGGTGA GCCCACCGAG GCCAGGGTGG TTCCGGGCCG GGGCATCTGC 1740
CTGGACCTCG CCATCCTGGA TAGTGCCTTC CTGCTGTCCC AGGTGGCCCC ATCCCTGTTT 1800
ATGGGCTCCA TTGTCCAGCT CAGCCAGTCT GTCACTGCCT ATATGGTGTC TGCCGCAGGC 1860
CTGGGTCTGG TCGCCATTTA CTTTGCTACA CAGGTAGTAT TTGACAAGAG CGACTTGGCC 1920
AAATACTCAG CGGGTGGACA CCATCACCAT CACCATTAA  1959
```

FIG. 6. Cloning strategy for generation of plasmid pRIT 15201



Rec'd PCT/PTO 08 DEC 2004 10/517420

FIG. 7. Plasmid map of pRIT15201

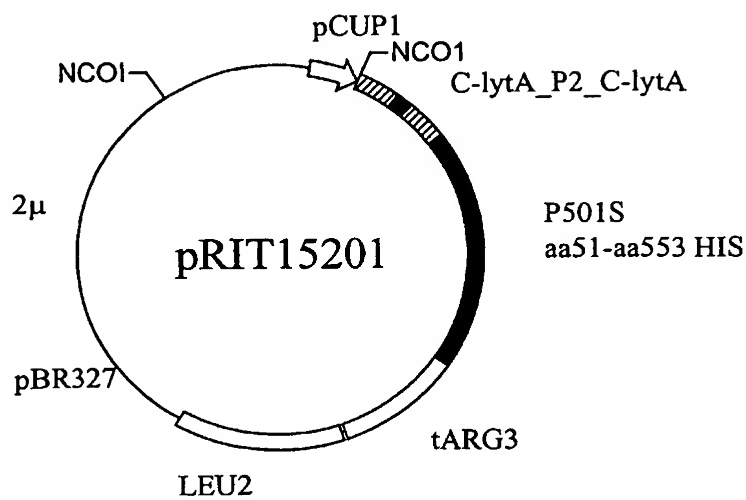
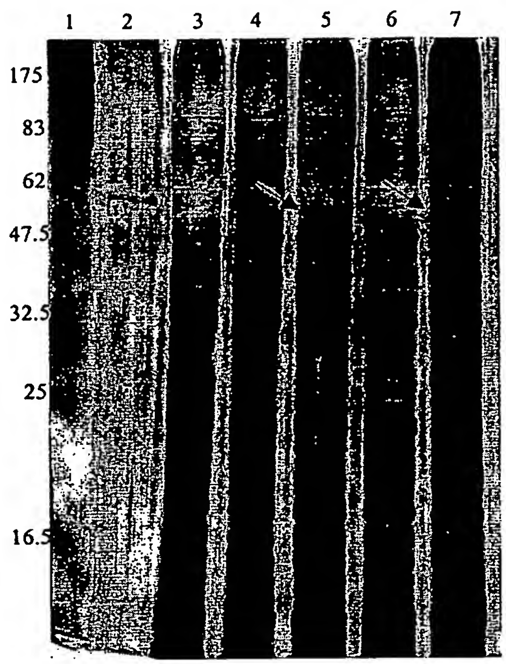
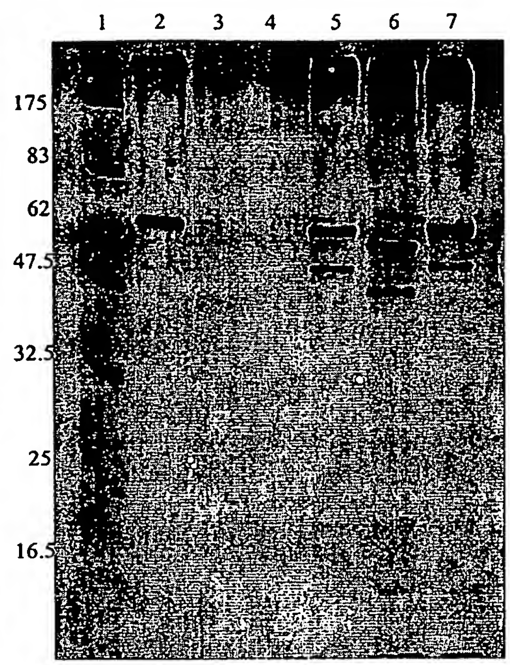


FIG. 8. Comparative expression of CPC P501 and P501 in *S.cerevisiae* strain DC5 (gel Laemmli 10%)



Silver staining



Western blot anti P501
(Monoclonal antibody)

1 MW Biolabs (175/83/62/47.5/32.5/16.5 Kda)
2 Y1796 purified
3 Y1795 Crude Extract (negative control)
4 SC333 Crude Extract
5 Y1796 Crude Extract
6 Y1790 Crude Extract
7 Y1802 Crude Extract

FIG. 9A.

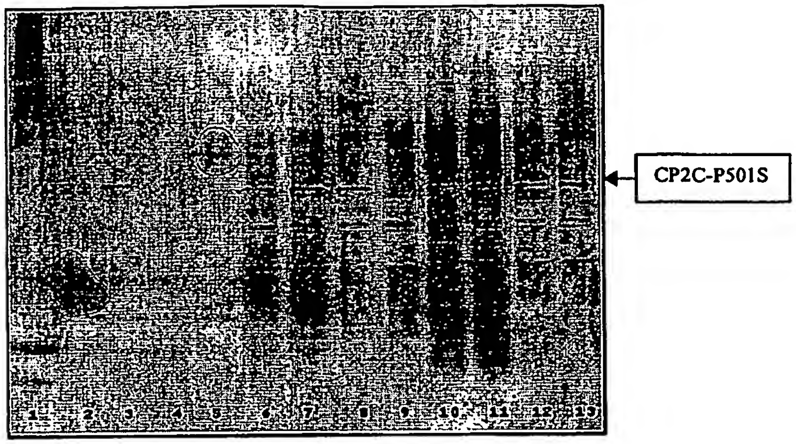
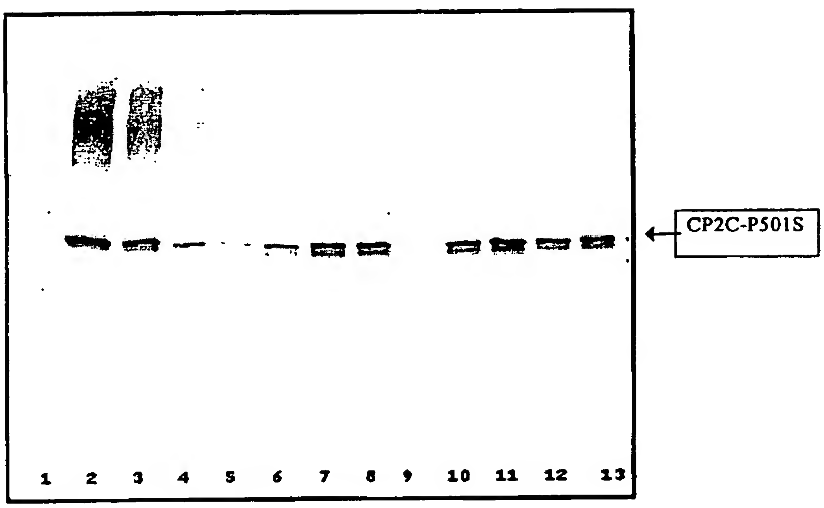


FIG. 9B.



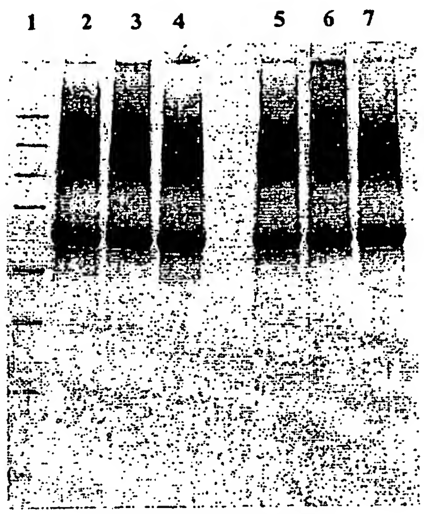
1 - Molecular Weight Marker (Biolabs - Grow Range)175; 83; 62; 47.5; 32.5; 25; 16.5; 6.5 kD - 10
2 - Purified Reference CP2CP501S/12 135 ng
3 - Purified Reference CP2CP501S/12 67.8 ng
4 - Purified Reference CP2CP501S/12 33.9 ng
5 - Purified Reference CP2CP501S/12 16.9 ng
6 - Fermentation PRO119-21h30
7 - Fermentation PRO124-21h30
8 - Fermentation PRO124-22h30
9 - Fermentation PRO127-0 h
10 - Fermentation PRO127-4 h
11 - Fermentation PRO127-6 h
12 - Fermentation PRO127-22h20
13 - Fermentation PRO127-22h45

FIG. 10. Purification scheme of CPC-P501-His produced by Y1796.

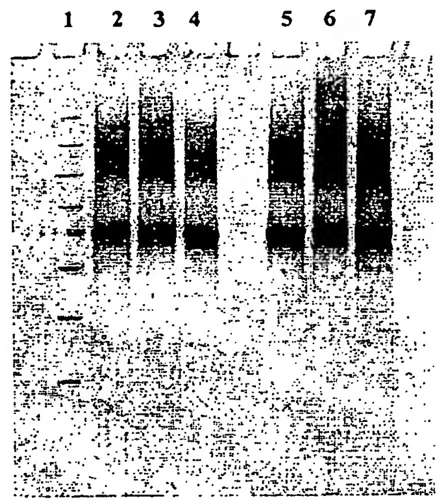
S. Cerevisiae cells	
↓	
Dyno-mill disruption	OD 120 / 2 passes / 20 mM Tris pH 8.5 - 5 mM EDTA
↓	
Centrifugation	12.000 g / RT / 90 min (supernatant discarded)
↓	
Pellet washing step 1	20 mM Tris pH 8.5 - 0.15 M NaCl - 2.0 M Guanidine.HCl - 0.1% Empigen (30 min / RT)
↓	
Centrifugation	12.000 g / RT / 60 min (supernatant discarded)
↓	
Pellet washing step 2	20 mM Tris pH 8.5 - 0.15 M NaCl - 4.0 M Urea
↓	
Centrifugation	12.000 g / RT / 30 min (supernatant discarded)
↓	
Solubilisation / Reduction	20 mM Tris pH 8.5 - 0.15 M NaCl - 8.0 M Urea - 1% SDS - 0.2 M Glutathion (60 min / RT)
↓	
Centrifugation	12.000 g / RT / 30 min (pellet discarded)
↓	
Carbamidomethylation	0.3 M Iodoacetamide (30 min / RT / in the dark) / pH adjusted to 8.5 (with 5 M NaOH solution) before incubation
↓	
R/C Supernatant	
↓	
10-fold dilution and pH adjustment (8.5)	Dilution buffer: 20 mM Tris pH 8.5 - 1 M NaCl - 8.0 M Urea
↓	
Immobilised metal ion affinity chromatography on Ni^{++}-Chelating Sepharose FF (Amersham) (10x25 cm column – 2000 ml)	Equilibration buffer: 20 mM Tris pH 8.5 - 0.9 M NaCl - 8.0 M Urea - 0.1% SDS Washing buffers: 1) Equilibration buffer 2) 20 mM Tris pH 8.5 - 0.15 M NaCl - 8.0 M Urea - 0.1% SDS 3) 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80

	Elution buffer: 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80 - 0.5 M Imidazole
↓	
2-fold dilution and pH adjustment (10.0)	20 mM Piperazine pH 10.0 - 8.0 M Urea - 0.1% Tween 80
↓	
Anion exchange chromatography on Q Sepharose FF (Amersham) (2,6 x 6.5 cm column - 35 ml)	Equilibration buffer: 20 mM Piperazine pH 10.0 - 8.0 M Urea - 0.1% Tween 80 Washing buffers: 1) Equilibration buffer 2) 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80 Elution buffer: 20 mM Tris pH 7.5 - 8.0 M Urea - 0.1% Tween 80 - 0.5 M NaCl
↓	
Concentration/Diafiltration (Pall - Omega 10 kDa - 200 cm²)	+/- 3-fold concentration Diafiltration buffer: Tris 20 mM pH 7.5
↓	
Sterile filtration (Millipore - Millex GV 0.22µm)	
↓	
Purified bulk	Final buffer: 20 mM Tris pH 7.5 - +/- 0.3% Tween 80
↓	
Storage –20°C	

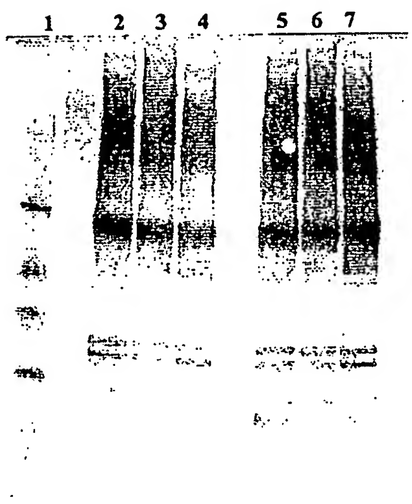
FIG. 11. Pattern of CPC P501 His purified protein (4-12% Novex Nu-Page polyacrylamide precasted gels)



Coomassie Blue R250



Daiichi Silver Staining



Western Blot anti P501S
(Monoclonal antibody)

1: MW (250/150/75/50/37/25/15/10 kDa)
2: Purified bulk A (reducing conditions)
3: Purified bulk B (reducing conditions)
4: Purified bulk C (reducing conditions)
5: Purified bulk A (non reducing conditions)
6: Purified bulk B (non reducing conditions)
7: Purified bulk C (non reducing conditions)

BEST AVAILABLE COPY

FIG. 12. Native full-length P501S sequence (SEQ ID NO:17 & 43)

Nucleotide sequence: SEQ ID NO.17

Polypeptide sequence: SEQ ID NO.43

```
          ######
          GCCACCATGGTCCAGAGGCTGTGGGTGAGCCGCCTGCTGCGGCACCGG
                M  V  Q  R  L  W  V  S  R  L  L  R  H  R   14

AAAGCCCAGCTCTTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTTGGCCGCA
K  A  Q  L  L  L  V  N  L  L  T  F  G  L  E  V  C  L  A  A   34

GGCATCACCTATGTGCCGCCTCTGCTGCTGGAAGTGGGGGTAGAGGAGAAGTTCATGACC
G  I  T  Y  V  P  P  L  L  L  E  V  G  V  E  E  K  F  M  T   54

ATGGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCC
M  V  L  G  I  G  P  V  L  G  L  V  C  V  P  L  L  G  S  A   74

AGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGGCCCTTCATCTGGGCACTGTCCTTG
S  D  H  W  R  G  R  Y  G  R  R  R  P  F  I  W  A  L  S  L   94

GGCATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGC
G  I  L  L  S  L  F  L  I  P  R  A  G  W  L  A  G  L  L  C   114

CCGGATCCCAGGCCCCTGGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACTTC
P  D  P  R  P  L  E  L  A  L  L  I  L  G  V  G  L  L  D  F   134

TGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGGACCCG
C  G  Q  V  C  F  T  P  L  E  A  L  L  S  D  L  F  R  D  P   154

GACCACTGTCGCCAGGCCTACTCTGTCTATGCCTTCATGATCAGTCTTGGGGGCTGCCTG
D  H  C  R  Q  A  Y  S  V  Y  A  F  M  I  S  L  G  G  C  L   174

GGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCCCTACCTGGGCACC
G  Y  L  L  P  A  I  D  W  D  T  S  A  L  A  P  Y  L  G  T   194

CAGGAGGAGTGCCTCTTTGGCCTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACA
Q  E  E  C  L  F  G  L  L  T  L  I  F  L  T  C  V  A  A  T   214

CTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCGGCC
L  L  V  A  E  E  A  A  L  G  P  T  E  P  A  E  G  L  S  A   234

CCCTCCTTGTCGCCCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGC
P  S  L  S  P  H  C  C  P  C  R  A  R  L  A  F  R  N  L  G   254

GCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCCGCATGCCCCGCACCCTGCGCCGGCTC
A  L  L  P  R  L  H  Q  L  C  C  R  M  P  R  T  L  R  R  L   274

TTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCTGTTTTACACGGAT
F  V  A  E  L  C  S  W  M  A  L  M  T  F  T  L  F  Y  T  D   294

TTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGG
F  V  G  E  G  L  Y  Q  G  V  P  R  A  E  P  G  T  E  A  R   314

AGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATC
R  H  Y  D  E  G  V  R  M  G  S  L  G  L  F  L  Q  C  A  I   334

TCCCTGGTCTTCTCTCTGGTCATGGACCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTC
S  L  V  F  S  L  V  M  D  R  L  V  Q  R  F  G  T  R  A  V   354
```

TATTTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCCACAGT
Y L A S V A A F P V A A G A T C L S H S 374

GTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCAGATC
V A V V T A S A A L T G F T F S A L Q I 394

CTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCAAATAC
L P Y T L A S L Y H R E K Q V F L P K Y 414

CGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGC
R G D T G G A S S E D S L M T S F L P G 434

CCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTC
P K P G A P F P N G H V G A G G S G L L 454

CCACCTCCACCCGCGCTCTGCGGGGCCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGTG
P P P P A L C G A S A C D V S V R V V V 474

GGTGAGCCCACCGAGGCCAGGGTGGTTCCGGGCCGGGGCATCTGCCTGGACCTCGCCATC
G E P T E A R V V P G R G I C L D L A I 494

CTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTC
L D S A F L L S Q V A P S L F M G S I V 514

CAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGGGTCTGGTCGCC
Q L S Q S V T A Y M V S A A G L G L V A 534

ATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAG
I Y F A T Q V V F D K S D L A K Y S A * 554

GTCGAG

FIG. 13. Sequence of the CPC-P501S expression cassette of JNW735 (SEQ ID NO:18 & 44)

Nucleotide sequence: SEQ ID NO.18

Polypeptide sequence: SEQ ID NO.44

```
######
GCCACCATGGCGGCCGCTTACGTACATTCCGACGGCTCTTATCCAAAA
      M  A  A  A  Y  V  H  S  D  G  S  Y  P  K   14

GACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTCAGGCTATATGCTT
D  K  F  E  K  I  N  G  T  W  Y  Y  F  D  S  S  G  Y  M  L   34

GCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAA
A  D  R  W  R  K  H  T  D  G  N  W  Y  W  F  D  N  S  G  E   54

ATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCC
M  A  T  G  W  K  K  I  A  D  K  W  Y  Y  F  N  E  E  G  A   74

ATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGC
M  K  T  G  W  V  K  Y  K  D  T  W  Y  Y  L  D  A  K  E  G   94

GCCATGCAATACATCAAGGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTA
A  M  Q  Y  I  K  A  N  S  K  F  I  G  I  T  E  G  V  M  V   114

TCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGA
S  N  A  F  I  Q  S  A  D  G  T  G  W  Y  Y  L  K  P  D  G   134

ACACTGGCAGACAGGCCAGAAAAGTTCATGTACATGGTGCTGGGCATTGGTCCAGTGCTG
T  L  A  D  R  P  E  K  F  M  Y  M  V  L  G  I  G  P  V  L   154

GGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGC
G  L  V  C  V  P  L  L  G  S  A  S  D  H  W  R  G  R  Y  G   174

CGCCGCCGGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTTCTCATC
R  R  R  P  F  I  W  A  L  S  L  G  I  L  L  S  L  F  L  I   194

CCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCCTGGAGCTGGCA
P  R  A  G  W  L  A  G  L  L  C  P  D  P  R  P  L  E  L  A   214

CTGCTCATCCTGGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTG
L  L  I  L  G  V  G  L  L  D  F  C  G  Q  V  C  F  T  P  L   234

GAGGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTCGCCAGGCCTACTCTGTC
E  A  L  L  S  D  L  F  R  D  P  D  H  C  R  Q  A  Y  S  V   254

TATGCCTTCATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGG
Y  A  F  M  I  S  L  G  G  C  L  G  Y  L  L  P  A  I  D  W   274

GACACCAGTGCCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCTCTTTGGCCTGCTC
D  T  S  A  L  A  P  Y  L  G  T  Q  E  E  C  L  F  G  L  L   294

ACCCTCATCTTCCTCACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGCAGCGCTG
T  L  I  F  L  T  C  V  A  A  T  L  L  V  A  E  E  A  A  L   314

GGCCCCACCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCACTGCTGTCCA
G  P  T  E  P  A  E  G  L  S  A  P  S  L  S  P  H  C  C  P   334
```

```
TGCCGGGCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTG
C  R  A  R  L  A  F  R  N  L  G  A  L  L  P  R  L  H  Q  L    354

TGCTGCCGCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATG
C  C  R  M  P  R  T  L  R  R  L  F  V  A  E  L  C  S  W  M    374

GCACTCATGACCTTCACGCTGTTTTACACGGATTTCGTGGGCGAGGGGCTGTACCAGGGC
A  L  M  T  F  T  L  F  Y  T  D  F  V  G  E  G  L  Y  Q  G    394

GTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGATG
V  P  R  A  E  P  G  T  E  A  R  R  H  Y  D  E  G  V  R  M    414

GGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGGTCATGGAC
G  S  L  G  L  F  L  Q  C  A  I  S  L  V  F  S  L  V  M  D    434

CGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCT
R  L  V  Q  R  F  G  T  R  A  V  Y  L  A  S  V  A  A  F  P    454

GTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCC
V  A  A  G  A  T  C  L  S  H  S  V  A  V  V  T  A  S  A  A    474

CTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTAC
L  T  G  F  T  F  S  A  L  Q  I  L  P  Y  T  L  A  S  L  Y    494

CACCGGGAGAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGT
H  R  E  K  Q  V  F  L  P  K  Y  R  G  D  T  G  G  A  S  S    514

GAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAAT
E  D  S  L  M  T  S  F  L  P  G  P  K  P  G  A  P  F  P  N    534

GGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCGCGCTCTGCGGGGCC
G  H  V  G  A  G  G  S  G  L  L  P  P  P  P  A  L  C  G  A    554

TCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGGGTGAGCCCACCGAGGCCAGGGTGGTT
S  A  C  D  V  S  V  R  V  V  V  G  E  P  T  E  A  R  V  V    574

CCGGGCCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCCTGCTGTCCCAG
P  G  R  G  I  C  L  D  L  A  I  L  D  S  A  F  L  L  S  Q    594

GTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTAT
V  A  P  S  L  F  M  G  S  I  V  Q  L  S  Q  S  V  T  A  Y    614

ATGGTGTCTGCCGCAGGCCTGGGTCTGGTCGCCATTTACTTTGCTACACAGGTAGTATTT
M  V  S  A  A  G  L  G  L  V  A  I  Y  F  A  T  Q  V  V  F    634

GACAAGAGCGACTTGGCCAAATACTCAGCGTAGGTCGAG
D  K  S  D  L  A  K  Y  S  A  *                               645
```

FIG. 14 – Codon optimised P501S sequences (SEQ ID NO:19-20)

SEQ ID NO:19

ATGGTGCAGCGGCTCTGGGTGAGCCGCCTCCTGCGGCATCGCAAGGCCCAGCTCCTGCTGGTGAATCTGCTCA
CATTCGGCCTGGAGGTGTGCCTGGCCGCCGGCATCACCTACGTGCCCCCCCTCCTGCTGGAGGTGGGAGTCGA
GGAGAAGTTCATGACCATGGTGCTGGGCATTGGGCCCGTCCTGGGCCTCGTGTGCGTGCCTCTCCTCGGCAGC
GCTTCCGACCATTGGCGCGGCCGGTATGGCCGCAGGAGACCCTTCATCTGGGCTCTGAGTCTCGGCATCCTGC
TGAGCCTGTTCCTGATCCCTCGGGCCGGCTGGCTGGCCGGGCTGCTGTGCCCCGATCCTCGGCCCCTGGAGCT
GGCCCTGCTGATCCTCGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGCTTCACGCCCCTGGAGGCACTG
CTGAGCGACCTGTTCCGGGACCCCGACCATTGCCGCCAGGCGTACAGCGTGTACGCCTTCATGATCTCCCTGG
GAGGCTGCCTGGGCTACCTGCTCCCCGCCATCGATTGGGACACCAGCGCACTCGCCCCCTATCTCGGAACACA
GGAGGAATGCCTGTTCGGATTGTTGACGCTCATCTTCCTCACGTGCGTCGCGGCCACCCTGTTGGTGGCCGAG
GAGGCCGCCCTGGGGCCCACCGAGCCGGCCGAGGGACTGAGCGCCCCGAGCCTGAGTCCACACTGCTGCCCTT
GCCGGGCCCGCCTGGCCTTCCGTAATCTGGGCGCCCTCCTGCCTCGGCTCCATCAGCTGTGTTGCAGAATGCC
TAGGACGCTGCGGCGCCTGTTCGTCGCTGAGTTGTGCTCCTGGATGGCTCTCATGACCTTCACCCTGTTTTAT
ACGGACTTCGTCGGGGAGGGCCTGTACCAGGGGGTGCCGCGCGCCGAGCCCGGGACAGAGGCGCGCCGCCACT
ACGACGAGGGAGTGCGTATGGGCTCCCTGGGCCTCTTCTTGCAGTGCGCCATCAGTCTGGTTTTCTCTCTGGT
CATGGACAGGCTGGTGCAGCGCTTCGGAACCCGGGCGGTGTACCTGGCGAGCGTGGCCGCCTTCCCCGTGGCT
GCCGGCGCCACCTGCCTCTCTCACTCGGTGGCCGTGGTCACCGCCAGCGCCGCCCTGACCGGGTTCACCTTCT
CTGCCCTGCAGATTCTGCCTTACACCCTGGCCAGCCTGTACCATCGCGAGAAACAGGTGTTTCTCCCCAAGTA
CAGAGGCGACACCGGGGGCGCCTCCAGCGAGGACAGCCTCATGACCTCCTTCCTGCCTGGCCCCAAGCCCGGC
GCCCCTTTCCCCAACGGGCACGTGGGCGCCGGCGGGAGTGGGCTCCTGCCCCCCCCTCCTGCGCTGTGCGGGG
CCAGCGCCTGCGACGTGAGCGTGCGCGTGGTGGTGGGCGAGCCCACCGAGGCCCGCGTGGTGCCGGGCAGAGG
CATTTGTCTGGACCTGGCCATCCTCGACTCCGCCTTCCTCCTCAGCCAGGTGGCCCCGTCCCTCTTCATGGGC
TCTATCGTCCAGCTGTCTCAGAGCGTCACCGCTTACATGGTGTCCGCTGCTGGACTGGGCTTGGTGGCTATTT
ATTTCGCCACCCAGGTGGTGTTCGACAAGAGCGACCTGGCCAAATACTCCGCCTGA

SEQ ID NO:20

ATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGGCCCAGTTGCTGCTGGTGAACCTGCTGA
CTTTCGGACTGGAGGTGTGCCTGGCTGCCGGGATCACGTACGTGCCCCCCCTGCTGCTGGAGGTGGGCGTGGA
GGAGAAGTTCATGACAATGGTGCTGGGCATCGGCCCCGTCCTGGGCCTCGTGTGTGTGCCCCTCCTCGGGAGT
GCGTCCGATCATTGGCGGGGCCGCTACGGCCGCCGCAGACCGTTCATCTGGGCCCTGAGCCTGGGGATCCTGC
TCTCTCTCTTCCTGATCCCCCGGGCCGGCTGGCTGGCCGGCCTGCTGTGTCCCGACCCCCGCCCTCTGGAGCT
GGCCCTCCTGATCCTGGGCGTGGGCTTGTTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCCTGGAGGCTCTG
CTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGG

GGGGATGCCTGGGCTATCTGCTGCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTCA
GGAGGAGTGCCTGTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCCTGCTGGTGGCCGAG
GAGGCGGCCCTGGGGCCCACCGAGCCCGCCGAGGGCCTGAGCGCTCCCAGCCTGAGCCCCCATTGCTGCCCGT
GCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCCTGCATCAGCTGTGCTGTCGCATGCC
TCGCACCCTGCGCCGCCTGTTCGTCGCTGAGCTCTGTTCCTGGATGGCCCTGATGACGTTCACCCTCTTCTAC
ACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGCCCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATT
ACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCAGTCTGGTGTTCTCTCTGGT
GATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCCGTGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCGCC
GCCGGCGCGACCTGCCTGTCTCATTCTGTCGCCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCA
GTGCGCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCCTGCCCAAGTA
CCGCGGGGACACAGGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCCCCAAGCCGGGG
GCCCCTTTCCCCAACGGCCATGTCGGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCCTGTGCGGCG
CTAGTGCCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGGAGCCCACCGAGGCTAGGGTCGTGCCTGGCCGGGG
GATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCCTGCTCTCCCAGGTGGCGCCCAGCCTGTTCATGGGC
AGTATCGTGCAGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGGCCTGGGGTTGGTGGCCATCT
ACTTTGCCACCCAGGTCGTGTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGA

10/517420

FIG. 15 – Re-engineered codon optimised sequence 19 (SEQ ID NO:21)

GACGGCTAGCGCCACCATGGTGCAGCGGCTCTGGGTGAGCCGCCTCCTGCGGCATCGCAAGGCCCAGCTCCTG
CTGGTGAATCTGCTCACATTCGGCCTGGAGGTGTGCCTGGCCGCCGGCATCACCTACGTGCCCCCCCTCCTGC
TGGAGGTGGGAGTCGAGGAGAAGTTCATGACCATGGTGCTGGGCATTGGGCCCGTCCTGGGCCTCGTGTGCGT
GCCTCTCCTCGGCAGCGCTTCCGACCATTGGCGCGGCCGGTATGGCCGCAGGAGACCCTTCATCTGGGCTCTG
AGTCTCGGCATCCTGCTGAGCCTGTTCCTGATCCCTCGGGCCGGCTGGCTGGCCGGGCTGCTGTGCCCCGATC
CTCGGCCCCTGGAGCTGGCCCTGCTGATCCTCGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGCTTCAC
GCCCCTGGAGGCACTGCTGAGCGACCTGTTCCGGGACCCCGACCATTGCCGCCAGGCGTACAGCGTGTACGCC
TTCATGATCTCCCTGGGAGGCTGCCTGGGCTACCTGCTCCCCGCCATCGATTGGGACACCAGCGCACTCGCCC
CCTATCTCGGAACACAGGAGGAATGCCTGTTCGGA$\boxed{C}$TG$\boxed{C}$TGACGCTCATCTTCCTCACGTGCGTCGCGGCCAC
CCTGTTGGTGGCCGAGGAGGCCGCCCTGGGGCCCACCGAGCCGGCCGAGGGACTGAGCGCCCCGAGCCTGAGT
CCACACTGCTGCCCTTGCCGGGCCCGCCTGGCCTTCCGTAATCTGGGCGCCCTCCTGCCTCGGCTCCATCAGC
TGTGTTGCAGAATGCCTAGGACGCTGCGGCGCCTGTTCGTCGCTGAGTTGTGCTCCTGGATGGCTCTCATGAC
CTTCACCCTGTTTTATACGGACTTCGTCGGGGAGGGCCTGTACCAGGGGGTGCCGCGCGCCGAGCCCGGGACA
GAGGCGCGCCGCCACTACGACGAGGGAGTGCGTATGGGCTCCCTGGGCCTCTTCTTGCAGTGCGCCATCAGTC
TGGTTTTCTCTCTGGTCATGGACAGGCTGGTGCAGCGCTTCGGAACCCGGGCGGTGTACCTGGCGAGCGTGGC
CGCCTTCCCCGTGGCTGCCGGCGCCACCTGCCTCTCTCACTCGGTGGCCGTGGTCACCGCCAGCGCCGCCCTG
ACCGGGTTCACCTTCTCTGCCCTGCAGATTCTGCCTTACACCCTGGCCAGCCTGTACCATCGCGAGAAACAGG
TGTTTCTCCCCAAGTACAGAGGCGACACCGGGGGCGCCTCCAGCGAGGACAGCCTCATGACCTCCTTCCTGCC
TGGCCCCAAGCCCGGCGCCCCTTTCCCCAACGGGCACGTGGGCGCCGGCGGGAGTGGGCTCCTGCCCCCCCCT
CCTGCGCTGTGCGGGGCCAGCGCCTGCGACGTGAGCGTGCGCGTGGTGGTGGGCGAGCCCACCGAGGCCCGCG
TGGTGCCGGGCAGAGGCATTTGTCTGGACCTGGCCATCCTCGACTCCGCCTTCCTCCTCAGCCAGGTGGCCCC
GTCCCTCTTCATGGGCTCTATCGTCCAGCTGTCTCAGAGCGTCACCGCTTACATGGTGTCCGCTGCTGGACTG
GGCTTGGTGGCTATTTATTTCGCCACCCAGGTGGTGTTCGACAAGAGCGACCTGGCCAAATACTCCGCCTGAC
TCGAGGCAG

FIG. 16 – Re-engineered codon optimised sequence 20 (SEQ ID NO:22)

GACGGCTAGCGCCACCATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGGCCCAGTTGCTG
CTGGTGAACCTGCTGACTTTCGGACTGGAGGTGTGCCTGGCTGCCGGGATCACGTACGTGCCCCCCCTGCTGC
TGGAGGTGGGCGTGGAGGAGAAGTTCATGACAATGGTGCTGGGCATCGGCCCCGTCCTGGGCCTCGTGTGTGT
GCCCCTCCTCGGGAGTGCGTCCGATCATTGGCGGGGCCGCTACGGCCGCCGCAGACCGTTCATCTGGGCCCTG
AGCCTGGGCATCCTGCTCTCTCTCTTCCTGATCCCCCGGGCCGGCTGGCTGGCCGGCCTGCTGTGTCCCGACC
CCCGCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGTTTCAC
TCCCCTGGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGTACGCC
TTCATGATCAGTCTGGGGGGATGCCTGGGCTATCTGCTGCCCGCTATCGACTGGGACACCAGCGCCCTGGCCC
CCTACCTGGGGACTCAGGAGGAGTGCCTGTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCAC
CCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCCACCGAGCCCGCCGAGGGCCTGAGCGCTCCCAGCCTGAGC
CCCCATTGCTGCCCGTGCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCCTGCATCAGC
TGTGCTGTCGCATGCCTCGCACCCTGCGCCGCCTGTTCGTCGCTGAGCTCTGTTCCTGGATGGCCCTGATGAC
GTTCACCCTCTTCTACACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGCCCAGGGCCGAGCCCGGCACC
GAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCAGTC
TGGTGTTCTCTCTGGTGATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCCGTGTACCTCGCCTCTGTGGC
GGCTTTCCCCGTCGCCGCCGGCGCGACCTGCCTGTCTCATTCTGTCGCCGTGGTGACCGCCAGCGCCGCCCTG
ACCGGCTTCACCTTCAGTGCGCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGG
TGTTCCTGCCCAAGTACCGCGGGGACACAGGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCC
CGGCCCCAAGCCGGGGGCCCCTTTCCCCAACGGCCATGTCGGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCC
CCCGCCCTGTGCGGCGCTAGTGCCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGGAGCCCACCGAGGCTAGGG
TCGTGCCTGGCCGGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCCTGCTCTCCCAGGTGGCGCC
CAGCCTGTTCATGGGCAGTATCGTGCAGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGGCCTG
GGGTTGGTGGCCATCTACTTTGCCACCCAGGTCGTGTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGAC
TCGAGGCAG

FIG. 17 – The starting sequence for the optimisation of CPC (SEQ ID NO:23)

Four amino acids of P501S sequence are boxed.

ATGGCGGCCGCTTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGT
ACTACTTTGACAGTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTT
CGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGT
GCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGCAAT
ACATCAAGGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTTTATCCAGTCAGC
GGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAAAGTTCATGTAC

FIG. 18 – Representative codon optimised CPC sequences (SEQ ID NO:24-25)

SEQ ID NO:24

ATGGCCGCCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACAAGTTCGAGAAGATCAACGGGACATGGT
ACTACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTT
CGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGC
GCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCATGCAGT
ATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCAACGCCTTTATCCAGAGCGC
CGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCCTCGCGGATCGGCCCGAGAAGTTCATGTAC

SEQ ID NO:25

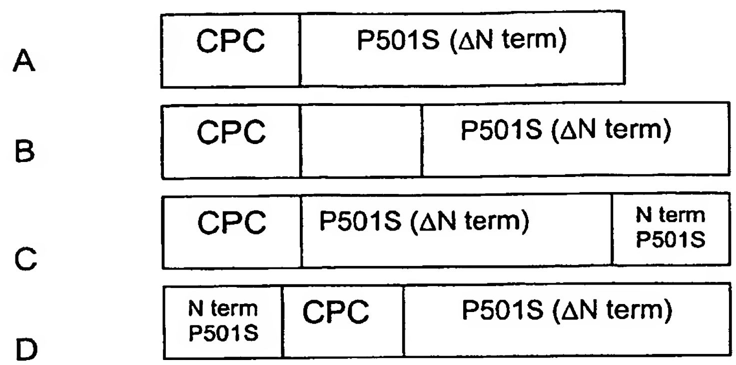
ATGGCCGCCGCCTACGTGCACAGCGACGGGTCCTACCCAAAGGACAAGTTCGAGAAGATCAACGGCACGTGGT
ACTATTTCGACAGCAGCGGCTACATGCTCGCCGATCGCTGGCGCAAGCACACCGACGGGAACTGGTACTGGTT
CGACAACTCTGGCGAGATGGCTACGGGGTGGAAGAAGATCGCCGACAAGTGGTACTACTTCAACGAGGAGGGC
GCCATGAAGACCGGGTGGGTGAAGTACAAGGACACCTGGTACTACCTGGACGCTAAGGAGGGCGCCATGCAGT
ACATCAAGGCCAACTCGAAGTTCATCGGGATCACCGAGGGCGTGATGGTCAGTAACGCTTTCATCCAGAGCGC
GGACGGCACAGGCTGGTATTACCTGAAGCCCGATGGCACCCTGGCGGACAGACCTGAGAAATTCATGTAC

FIG. 19 – Engineered CPC codon optimised sequence (SEQ ID NO:26)

SEQ ID NO:26

GACGGCTAGCGCCACCATGGCCGCCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACAAGTTCGAGAAG
ATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACG
GCAACTGGTACTGGTTCGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTA
TTTCAACGAGGAGGGCGCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAG
GAGGGCGCCATGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCAACG
CCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCCTCGCGGATCGGCCCGA
GAAGTTCATGTAC*TGA*CTCGAGGCAG

FIG. 20 – P501S CPC fusion candidate constructs and sequences



Construct A = SEQ ID NO:37 (nucleotide) & 45 (polypeptide)

```
GCGGCCGCGCCACCATGGCCGCCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
              M  A  A  A  Y  V  H  S  D  G  S  Y  P  K  D  K

AGTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
 F  E  K  I  N  G  T  W  Y  Y  F  D  S  S  G  Y  M  L  A  D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
 R  W  R  K  H  T  D  G  N  W  Y  W  F  D  N  S  G  E  M  A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGCGCCATGA
 T  G  W  K  K  I  A  D  K  W  Y  Y  F  N  E  E  G  A  M  K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
 T  G  W  V  K  Y  K  D  T  W  Y  Y  L  D  A  K  E  G  A  M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
 Q  Y  I  K  A  N  S  K  F  I  G  I  T  E  G  V  M  V  S  N

ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
 A  F  I  Q  S  A  D  G  T  G  W  Y  Y  L  K  P  D  G  T  L

TCGCGGATCGGCCCGAGAAGTTCATGTACATGGTGCTGGGCATCGGCCCCGTCCTGGGCC
 A  D  R  P  E  K  F  M  Y  M  V  L  G  I  G  P  V  L  G  L

TCGTGTGTGTGCCCCTCCTCGGGAGTGCGTCCGATCATTGGCGGGGCCGCTACGGCCGCC
 V  C  V  P  L  L  G  S  A  S  D  H  W  R  G  R  Y  G  R  R

GCAGACCGTTCATCTGGGCCCTGAGCCTGGGCATCCTGCTCTCTCTCTTCCTGATCCCCC
 R  P  F  I  W  A  L  S  L  G  I  L  L  S  L  F  L  I  P  R

GGGCCGGCTGGCTGGCCGGCCTGCTGTGTCCCGACCCCCGCCCTCTGGAGCTGGCCCTCC
 A  G  W  L  A  G  L  L  C  P  D  P  R  P  L  E  L  A  L  L

TGATCCTGGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCCTGGAGG
```

```
 I  L  G  V  G  L  L  D  F  C  G  Q  V  C  F  T  P  L  E  A
CTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGTACG
 L  L  S  D  L  F  R  D  P  D  H  C  R  Q  A  Y  S  V  Y  A
CCTTCATGATCAGTCTGGGGGATGCCTGGGCTATCTGCTGCCCGCTATCGACTGGGACA
 F  M  I  S  L  G  G  C  L  G  Y  L  L  P  A  I  D  W  D  T
CCAGCGCCCTGGCCCCCTACCTGGGGACTCAGGAGGAGTGCCTGTTCGGCCTGCTCACCT
 S  A  L  A  P  Y  L  G  T  Q  E  E  C  L  F  G  L  L  T  L
TGATCTTCCTGACGTGCGTCGCCGCCACCCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGC
 I  F  L  T  C  V  A  A  T  L  L  V  A  E  E  A  A  L  G  P
CCACCGAGCCCGCCGAGGGCCTGAGCGCTCCCAGCCTGAGCCCCCATTGCTGCCCGTGCA
 T  E  P  A  E  G  L  S  A  P  S  L  S  P  H  C  C  P  C  R
GGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCCTGCATCAGCTGTGCT
 A  R  L  A  F  R  N  L  G  A  L  L  P  R  L  H  Q  L  C  C
GTCGCATGCCTCGCACCCTGCGCCGCCTGTTCGTCGCTGAGCTCTGTTCCTGGATGGCCC
 R  M  P  R  T  L  R  R  L  F  V  A  E  L  C  S  W  M  A  L
TGATGACGTTCACCCTCTTCTACACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGC
 M  T  F  T  L  F  Y  T  D  F  V  G  E  G  L  Y  Q  G  V  P
CCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCT
 R  A  E  P  G  T  E  A  R  R  H  Y  D  E  G  V  R  M  G  S
CTCTGGGCCTCTTCCTGCAGTGCGCCATCAGTCTGGTGTTCTCTCTGGTGATGGACCGGC
 L  G  L  F  L  Q  C  A  I  S  L  V  F  S  L  V  M  D  R  L
TGGTGCAGCGCTTCGGCACCCGGGCCGTGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCG
 V  Q  R  F  G  T  R  A  V  Y  L  A  S  V  A  A  F  P  V  A
CCGCCGGCGCGACCTGCCTGTCTCATTCTGTCGCCGTGGTGACCGCCAGCGCCGCCCTGA
 A  G  A  T  C  L  S  H  S  V  A  V  V  T  A  S  A  A  L  T
CCGGCTTCACCTTCAGTGCGCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATC
 G  F  T  F  S  A  L  Q  I  L  P  Y  T  L  A  S  L  Y  H  R
GCGAGAAGCAGGTGTTCCTGCCCAAGTACCGCGGGGACACAGGGGGAGCTTCCTCTGAGG
 E  K  Q  V  F  L  P  K  Y  R  G  D  T  G  G  A  S  S  E  D
ACAGCCTGATGACCAGCTTCTTGCCCGGCCCCAAGCCGGGGGCCCCTTTCCCCAACGGCC
 S  L  M  T  S  F  L  P  G  P  K  P  G  A  P  F  P  N  G  H
ATGTCGGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCCTGTGCGGCGCTAGTG
 V  G  A  G  G  S  G  L  L  P  P  P  P  A  L  C  G  A  S  A
CCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGGAGCCCACCGAGGCTAGGGTCGTGCCTG
 C  D  V  S  V  R  V  V  V  G  E  P  T  E  A  R  V  V  P  G
GCCGGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCCTGCTCTCCCAGGTGG
 R  G  I  C  L  D  L  A  I  L  D  S  A  F  L  L  S  Q  V  A
CGCCCAGCCTGTTCATGGGCAGTATCGTGCAGCTGAGCCAGAGCGTGACCGCCTACATGG
 P  S  L  F  M  G  S  I  V  Q  L  S  Q  S  V  T  A  Y  M  V
```

```
TGAGCGCCGCCGGCCTGGGGTTGGTGGCCATCTACTTTGCCACCCAGGTCGTGTTCGACA
 S  A  A  G  L  G  L  V  A  I  Y  F  A  T  Q  V  V  F  D  K

AGAGCGATCTCGCCAAGTATAGCGCCTGAGGATCC
 S  D  L  A  K  Y  S  A  *
```

Construct B = SEQ ID NO:38 (nucleotide) & 46 (polypeptide)

```
GCGGCCGCGCCACCATGGCCGCCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
              M  A  A  A  Y  V  H  S  D  G  S  Y  P  K  D  K

AGTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
 F  E  K  I  N  G  T  W  Y  Y  F  D  S  S  G  Y  M  L  A  D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
 R  W  R  K  H  T  D  G  N  W  Y  W  F  D  N  S  G  E  M  A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGCGCCATGA
 T  G  W  K  K  I  A  D  K  W  Y  Y  F  N  E  E  G  A  M  K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
 T  G  W  V  K  Y  K  D  T  W  Y  Y  L  D  A  K  E  G  A  M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
 Q  Y  I  K  A  N  S  K  F  I  G  I  T  E  G  V  M  V  S  N

ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
 A  F  I  Q  S  A  D  G  T  G  W  Y  Y  L  K  P  D  G  T  L

TCGCGGATCGGCCCGAGATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAA
 A  D  R  P  E  M  V  Q  R  L  W  V  S  R  L  L  R  H  R  K

AGGCCCAGTTGCTGCTGGTGAACCTGCTGACTTTCGGACTGGAGGTGTGCCTGGCTGCCG
 A  Q  L  L  L  V  N  L  L  T  F  G  L  E  V  C  L  A  A  G

GGATCACGTACGTGCCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGAAGTTCATGACAA
 I  T  Y  V  P  P  L  L  L  E  V  G  V  E  E  K  F  M  T  M

TGGTGCTGGGCATCGGCCCCGTCCTGGGCCTCGTGTGTGTGCCCCTCCTCGGGAGTGCGT
 V  L  G  I  G  P  V  L  G  L  V  C  V  P  L  L  G  S  A  S

CCGATCATTGGCGGGGCCGCTACGGCCGCCGCAGACCGTTCATCTGGGCCCTGAGCCTGG
 D  H  W  R  G  R  Y  G  R  R  R  P  F  I  W  A  L  S  L  G

GCATCCTGCTCTCTCTCTTCCTGATCCCCCGGGCCGGCTGGCTGGCCGGCCTGCTGTGTC
 I  L  L  S  L  F  L  I  P  R  A  G  W  L  A  G  L  L  C  P

CCGACCCCCGCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCCTGCTGGACTTCT
 D  P  R  P  L  E  L  A  L  L  I  L  G  V  G  L  L  D  F  C

GCGGCCAGGTGTGTTTCACTCCCCTGGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCG
 G  Q  V  C  F  T  P  L  E  A  L  L  S  D  L  F  R  D  P  D

ACCACTGTAGGCAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGGGGGGATGCCTGG
 H  C  R  Q  A  Y  S  V  Y  A  F  M  I  S  L  G  G  C  L  G

GCTATCTGCTGCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTC
 Y  L  L  P  A  I  D  W  D  T  S  A  L  A  P  Y  L  G  T  Q
```

AGGAGGAGTGCCTGTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCC
E E C L F G L L T L I F L T C V A A T L

TGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCCACCGAGCCCGCCGAGGGCCTGAGCGCTC
L V A E E A A L G P T E P A E G L S A P

CCAGCCTGAGCCCCCATTGCTGCCCGTGCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCG
S L S P H C C P C R A R L A F R N L G A

CTTTGCTGCCCCGCCTGCATCAGCTGTGCTGTCGCATGCCTCGCACCCTGCGCCGCCTGT
L L P R L H Q L C C R M P R T L R R L F

TCGTCGCTGAGCTCTGTTCCTGGATGGCCCTGATGACGTTCACCCTCTTCTACACCGACT
V A E L C S W M A L M T F T L F Y T D F

TCGTGGGGGAGGGCCTGTACCAGGGCGTGCCCAGGGCCGAGCCCGGCACCGAGGCTAGGC
V G E G L Y Q G V P R A E P G T E A R R

GCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCA
H Y D E G V R M G S L G L F L Q C A I S

GTCTGGTGTTCTCTCTGGTGATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCCGTGT
L V F S L V M D R L V Q R F G T R A V Y

ACCTCGCCTCTGTGGCGGCTTTCCCCGTCGCCGCCGGCGCGACCTGCCTGTCTCATTCTG
L A S V A A F P V A A G A T C L S H S V

TCGCCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCAGTGCGCTCCAGATTC
A V V T A S A A L T G F T F S A L Q I L

TGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCCTGCCCAAGTACC
P Y T L A S L Y H R E K Q V F L P K Y R

GCGGGGACACAGGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCC
G D T G G A S S E D S L M T S F L P G P

CCAAGCCGGGGGCCCCTTTCCCCAACGGCCATGTCGGGGCGGGCGGCAGCGGCCTGCTCC
K P G A P F P N G H V G A G G S G L L P

CTCCCCCCCCCGCCCTGTGCGGCGCTAGTGCCTGCGACGTGAGCGTGCGGGTGGTGGTGG
P P P A L C G A S A C D V S V R V V V G

GGGAGCCCACCGAGGCTAGGGTCGTGCCTGGCCGGGGGATCTGCCTGGACCTGGCCATCC
E P T E A R V V P G R G I C L D L A I L

TCGACTCCGCCTTCCTGCTCTCCCAGGTGGCGCCCAGCCTGTTCATGGGCAGTATCGTGC
D S A F L L S Q V A P S L F M G S I V Q

AGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGGCCTGGGGTTGGTGGCCA
L S Q S V T A Y M V S A A G L G L V A I

TCTACTTTGCCACCCAGGTCGTGTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGAG
Y F A T Q V V F D K S D L A K Y S A *

GATCC

Construct C = SEQ ID NO:39 (nucleotide) & 47 (polypeptide)

```
GCGGCCGCGCCACCATGGCCGCCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
              M  A  A  A  Y  V  H  S  D  G  S  Y  P  K  D  K

AGTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
 F  E  K  I  N  G  T  W  Y  Y  F  D  S  S  G  Y  M  L  A  D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
 R  W  R  K  H  T  D  G  N  W  Y  W  F  D  N  S  G  E  M  A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGCGCCATGA
 T  G  W  K  K  I  A  D  K  W  Y  Y  F  N  E  E  G  A  M  K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
 T  G  W  V  K  Y  K  D  T  W  Y  Y  L  D  A  K  E  G  A  M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
 Q  Y  I  K  A  N  S  K  F  I  G  I  T  E  G  V  M  V  S  N

ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
 A  F  I  Q  S  A  D  G  T  G  W  Y  Y  L  K  P  D  G  T  L

TCGCGGATCGGCCCGAGAAGTTCATGTACATGGTGCTGGGCATCGGCCCCGTCCTGGGCC
 A  D  R  P  E  K  F  M  Y  M  V  L  G  I  G  P  V  L  G  L

TCGTGTGTGTGCCCCTCCTCGGGAGTGCGTCCGATCATTGGCGGGGCCGCTACGGCCGCC
 V  C  V  P  L  L  G  S  A  S  D  H  W  R  G  R  Y  G  R  R

GCAGACCGTTCATCTGGGCCCTGAGCCTGGGCATCCTGCTCTCTCTCTTCCTGATCCCCC
 R  P  F  I  W  A  L  S  L  G  I  L  L  S  L  F  L  I  P  R

GGGCCGGCTGGCTGGCCGGCCTGCTGTGTCCCGACCCCCGCCCTCTGGAGCTGGCCCTCC
 A  G  W  L  A  G  L  L  C  P  D  P  R  P  L  E  L  A  L  L

TGATCCTGGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCCTGGAGG
 I  L  G  V  G  L  L  D  F  C  G  Q  V  C  F  T  P  L  E  A

CTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGTACG
 L  L  S  D  L  F  R  D  P  D  H  C  R  Q  A  Y  S  V  Y  A

CCTTCATGATCAGTCTGGGGGGATGCCTGGGCTATCTGCTGCCCGCTATCGACTGGGACA
 F  M  I  S  L  G  G  C  L  G  Y  L  L  P  A  I  D  W  D  T

CCAGCGCCCTGGCCCCCTACCTGGGGACTCAGGAGGAGTGCCTGTTCGGCCTGCTCACCT
 S  A  L  A  P  Y  L  G  T  Q  E  E  C  L  F  G  L  L  T  L

TGATCTTCCTGACGTGCGTCGCCGCCACCCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGC
 I  F  L  T  C  V  A  A  T  L  L  V  A  E  E  A  A  L  G  P

CCACCGAGCCCGCCGAGGGCCTGAGCGCTCCCAGCCTGAGCCCCCATTGCTGCCCGTGCA
 T  E  P  A  E  G  L  S  A  P  S  L  S  P  H  C  C  P  C  R

GGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCCTGCATCAGCTGTGCT
 A  R  L  A  F  R  N  L  G  A  L  L  P  R  L  H  Q  L  C  C

GTCGCATGCCTCGCACCCTGCGCCGCCTGTTCGTCGCTGAGCTCTGTTCCTGGATGGCCC
 R  M  P  R  T  L  R  R  L  F  V  A  E  L  C  S  W  M  A  L

TGATGACGTTCACCCTCTTCTACACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGC
 M  T  F  T  L  F  Y  T  D  F  V  G  E  G  L  Y  Q  G  V  P
```

```
CCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCT
 R  A  E  P  G  T  E  A  R  R  H  Y  D  E  G  V  R  M  G  S

CTCTGGGCCTCTTCCTGCAGTGCGCCATCAGTCTGGTGTTCTCTCTGGTGATGGACCGGC
 L  G  L  F  L  Q  C  A  I  S  L  V  F  S  L  V  M  D  R  L

TGGTGCAGCGCTTCGGCACCCGGGCCGTGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCG
 V  Q  R  F  G  T  R  A  V  Y  L  A  S  V  A  A  F  P  V  A

CCGCCGGCGCGACCTGCCTGTCTCATTCTGTCGCCGTGGTGACCGCCAGCGCCGCCCTGA
 A  G  A  T  C  L  S  H  S  V  A  V  V  T  A  S  A  A  L  T

CCGGCTTCACCTTCAGTGCGCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATC
 G  F  T  F  S  A  L  Q  I  L  P  Y  T  L  A  S  L  Y  H  R

GCGAGAAGCAGGTGTTCCTGCCCAAGTACCGCGGGGACACAGGGGGAGCTTCCTCTGAGG
 E  K  Q  V  F  L  P  K  Y  R  G  D  T  G  G  A  S  S  E  D

ACAGCCTGATGACCAGCTTCTTGCCCGGCCCCAAGCCGGGGGCCCCTTTCCCCAACGGCC
 S  L  M  T  S  F  L  P  G  P  K  P  G  A  P  F  P  N  G  H

ATGTCGGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCCTGTGCGGCGCTAGTG
 V  G  A  G  G  S  G  L  L  P  P  P  P  A  L  C  G  A  S  A

CCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGGAGCCCACCGAGGCTAGGGTCGTGCCTG
 C  D  V  S  V  R  V  V  V  G  E  P  T  E  A  R  V  V  P  G

GCCGGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCCTGCTCTCCCAGGTGG
 R  G  I  C  L  D  L  A  I  L  D  S  A  F  L  L  S  Q  V  A

CGCCCAGCCTGTTCATGGGCAGTATCGTGCAGCTGAGCCAGAGCGTGACCGCCTACATGG
 P  S  L  F  M  G  S  I  V  Q  L  S  Q  S  V  T  A  Y  M  V

TGAGCGCCGCCGGCCTGGGGTTGGTGGCCATCTACTTTGCCACCCAGGTCGTGTTCGACA
 S  A  A  G  L  G  L  V  A  I  Y  F  A  T  Q  V  V  F  D  K

AGAGCGATCTCGCCAAGTATAGCGCCATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGC
 S  D  L  A  K  Y  S  A  M  V  Q  R  L  W  V  S  R  L  L  R

GCCATAGAAAGGCCCAGTTGCTGCTGGTGAACCTGCTGACTTTCGGACTGGAGGTGTGCC
 H  R  K  A  Q  L  L  L  V  N  L  L  T  F  G  L  E  V  C  L

TGGCTGCCGGGATCACGTACGTGCCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGTGAG
 A  A  G  I  T  Y  V  P  P  L  L  L  E  V  G  V  E  E  *

GATCC
```

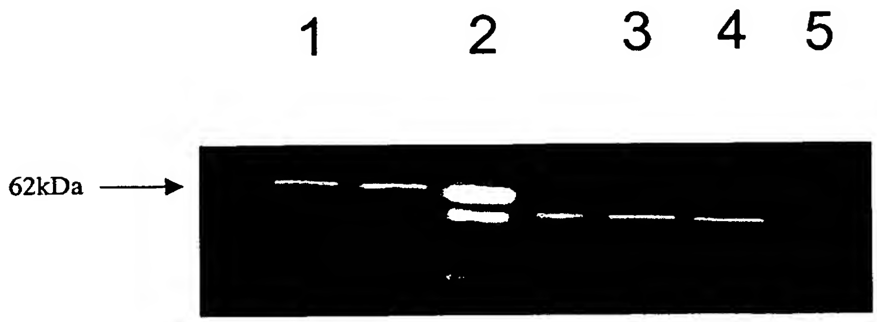
Construct D = SEQ ID NO:40 (nucleotide) & 48 (polypeptide)

```
GCGGCCGCGCCACCATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGG
                M  V  Q  R  L  W  V  S  R  L  L  R  H  R  K  A

CCCAGTTGCTGCTGGTGAACCTGCTGACTTTCGGACTGGAGGTGTGCCTGGCTGCCGGGA
 Q  L  L  L  V  N  L  L  T  F  G  L  E  V  C  L  A  A  G  I

TCACGTACGTGCCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGATGGCCGCCGCCTACG
 T  Y  V  P  P  L  L  L  E  V  G  V  E  E  M  A  A  A  Y  V
```

```
TGCATAGCGACGGGAGCTACCCCAAGGACAAGTTCGAGAAGATCAACGGGACATGGTACT
 H  S  D  G  S  Y  P  K  D  K  F  E  K  I  N  G  T  W  Y  Y

ACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACGGCAACT
 F  D  S  S  G  Y  M  L  A  D  R  W  R  K  H  T  D  G  N  W

GGTACTGGTTCGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGT
 Y  W  F  D  N  S  G  E  M  A  T  G  W  K  K  I  A  D  K  W

GGTACTATTTCAACGAGGAGGGCGCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCT
 Y  Y  F  N  E  E  G  A  M  K  T  G  W  V  K  Y  K  D  T  W

GGTACTACCTCGACGCCAAGGAGGGCGCCATGCAGTATATCAAGGCCAACAGCAAGTTCA
 Y  Y  L  D  A  K  E  G  A  M  Q  Y  I  K  A  N  S  K  F  I

TCGGCATCACCGAGGGAGTGATGGTCAGCAACGCCTTTATCCAGAGCGCCGACGGCACCG
 G  I  T  E  G  V  M  V  S  N  A  F  I  Q  S  A  D  G  T  G

GATGGTACTACTTGAAGCCGGACGGCACCCTCGCGGATCGGCCCGAGAAGTTCATGTACA
 W  Y  Y  L  K  P  D  G  T  L  A  D  R  P  E  K  F  M  Y  M

TGGTGCTGGGCATCGGCCCCGTCCTGGGCCTCGTGTGTGTGCCCCTCCTCGGGAGTGCGT
 V  L  G  I  G  P  V  L  G  L  V  C  V  P  L  L  G  S  A  S

CCGATCATTGGCGGGGCCGCTACGGCCGCCGCAGACCGTTCATCTGGGCCCTGAGCCTGG
 D  H  W  R  G  R  Y  G  R  R  R  P  F  I  W  A  L  S  L  G

GCATCCTGCTCTCTCTCTTCCTGATCCCCCGGGCCGGCTGGCTGGCCGGCCTGCTGTGTC
 I  L  L  S  L  F  L  I  P  R  A  G  W  L  A  G  L  L  C  P

CCGACCCCCGCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCCTGCTGGACTTCT
 D  P  R  P  L  E  L  A  L  L  I  L  G  V  G  L  L  D  F  C

GCGGCCAGGTGTGTTTCACTCCCCTGGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCG
 G  Q  V  C  F  T  P  L  E  A  L  L  S  D  L  F  R  D  P  D

ACCACTGTAGGCAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGGGGGGATGCCTGG
 H  C  R  Q  A  Y  S  V  Y  A  F  M  I  S  L  G  G  C  L  G

GCTATCTGCTGCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTC
 Y  L  L  P  A  I  D  W  D  T  S  A  L  A  P  Y  L  G  T  Q

AGGAGGAGTGCCTGTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCC
 E  E  C  L  F  G  L  L  T  L  I  F  L  T  C  V  A  A  T  L

TGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCCACCGAGCCCGCCGAGGGCCTGAGCGCTC
 L  V  A  E  E  A  A  L  G  P  T  E  P  A  E  G  L  S  A  P

CCAGCCTGAGCCCCCATTGCTGCCCGTGCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCG
 S  L  S  P  H  C  C  P  C  R  A  R  L  A  F  R  N  L  G  A

CTTTGCTGCCCCGCCTGCATCAGCTGTGCTGTCGCATGCCTCGCACCCTGCGCCGCCTGT
 L  L  P  R  L  H  Q  L  C  C  R  M  P  R  T  L  R  R  L  F

TCGTCGCTGAGCTCTGTTCCTGGATGGCCCTGATGACGTTCACCCTCTTCTACACCGACT
 V  A  E  L  C  S  W  M  A  L  M  T  F  T  L  F  Y  T  D  F

TCGTGGGGGAGGGCCTGTACCAGGGCGTGCCCAGGGCCGAGCCCGGCACCGAGGCTAGGC
```

```
 V  G  E  G  L  Y  Q  G  V  P  R  A  E  P  G  T  E  A  R  R

GCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCA
 H  Y  D  E  G  V  R  M  G  S  L  G  L  F  L  Q  C  A  I  S

GTCTGGTGTTCTCTCTGGTGATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCCGTGT
 L  V  F  S  L  V  M  D  R  L  V  Q  R  F  G  T  R  A  V  Y

ACCTCGCCTCTGTGGCGGCTTTCCCCGTCGCCGCCGGCGCGACCTGCCTGTCTCATTCTG
 L  A  S  V  A  A  F  P  V  A  A  G  A  T  C  L  S  H  S  V

TCGCCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCAGTGCGCTCCAGATTC
 A  V  V  T  A  S  A  A  L  T  G  F  T  F  S  A  L  Q  I  L

TGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCCTGCCCAAGTACC
 P  Y  T  L  A  S  L  Y  H  R  E  K  Q  V  F  L  P  K  Y  R

GCGGGGACACAGGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCC
 G  D  T  G  G  A  S  S  E  D  S  L  M  T  S  F  L  P  G  P

CCAAGCCGGGGGCCCCTTTCCCCAACGGCCATGTCGGGGCGGGCGGCAGCGGCCTGCTCC
 K  P  G  A  P  F  P  N  G  H  V  G  A  G  G  S  G  L  L  P

CTCCCCCCCCCGCCCTGTGCGGCGCTAGTGCCTGCGACGTGAGCGTGCGGGTGGTGGTGG
 P  P  P· A  L  C  G  A  S  A  C  D  V  S  V  R  V  V  V  G

GGGAGCCCACCGAGGCTAGGGTCGTGCCTGGCCGGGGGATCTGCCTGGACCTGGCCATCC
 E  P  T  E  A  R  V  V  P  G  R  G  I  C  L  D  L  A  I  L

TCGACTCCGCCTTCCTGCTCTCCCAGGTGGCGCCCAGCCTGTTCATGGGCAGTATCGTGC
 D  S  A  F  L  L  S  Q  V  A  P  S  L  F  M  G  S  I  V  Q

AGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGGCCTGGGGTTGGTGGCCA
 L  S  Q  S  V  T  A  Y  M  V  S  A  A  G  L  G  L  V  A  I

TCTACTTTGCCACCCAGGTCGTGTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGAG
 Y  F  A  T  Q  V  V  F  D  K  S  D  L  A  K  Y  S  A  *

GATCC
```

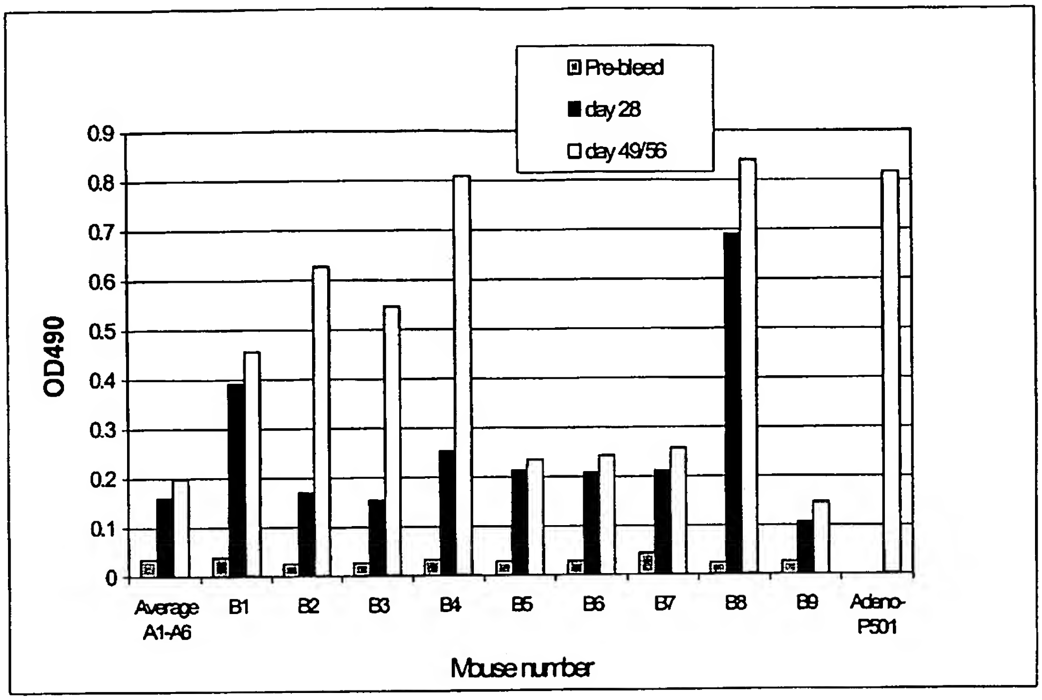
FIG. 21 – Western blot analysis of CHO cells following transient transfection with P501S (JNW680), CPC-P501S (JNW735) and empty vector control.



Lane	Sample
1	CPC-P501S (JNW735)
2	CPC P501S protein (62.5ng)
3	P501S (JNW680)
4	P501S (JNW680)
5	Empty vector control

10/517420

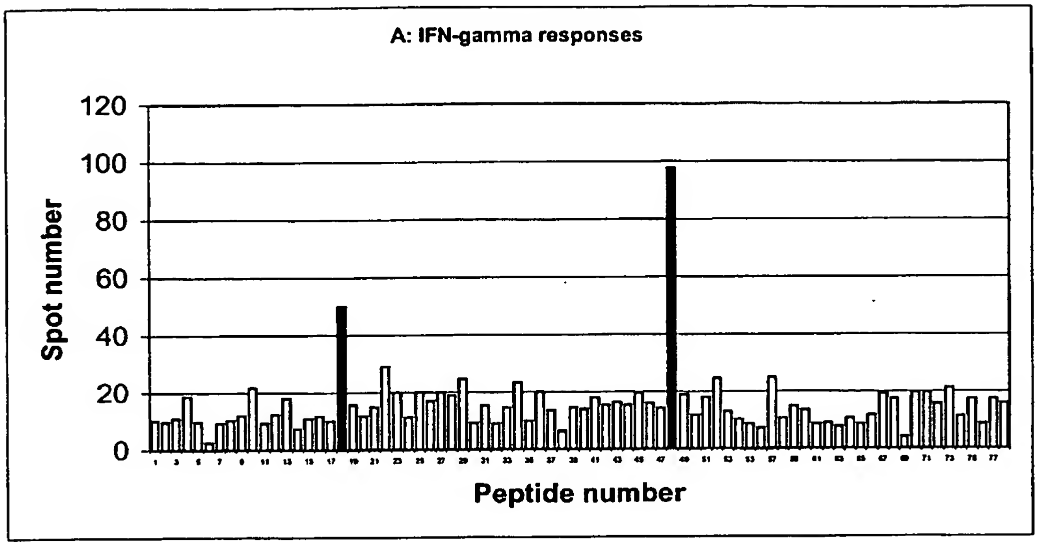
FIG. 22 – Anti-P501S antibody responses following immunisation at day0, 21 & 42 with pVAC-P501S (JNW680, mice B1-9) or Empty vector (pVAC, mice A1-6).

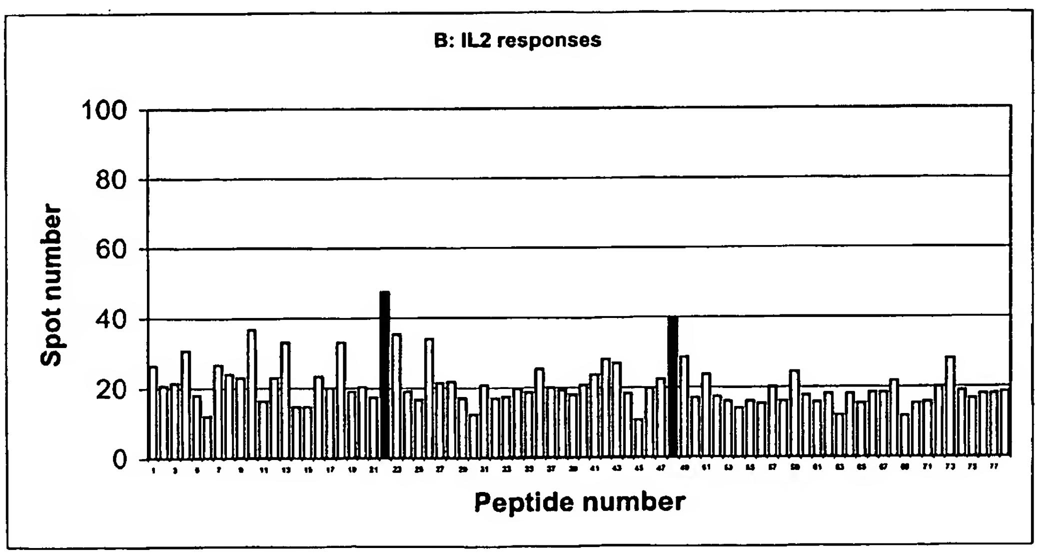


Rec'd PCT/PTO 08 DEC 2004

10/517420

FIG. 23 – Peptide library screen using C57BL/6 mice immunised at day 0, 21, 42, and 70 with pVAC-P501S (JNW680).



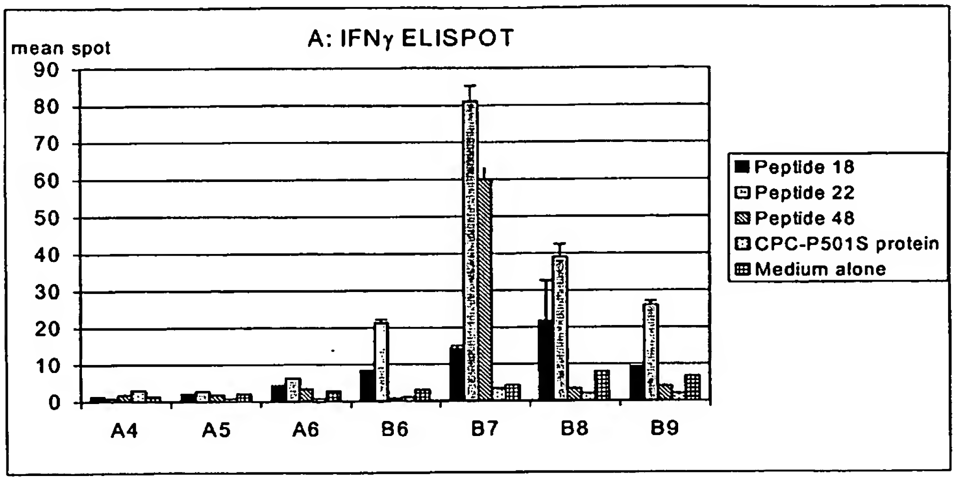


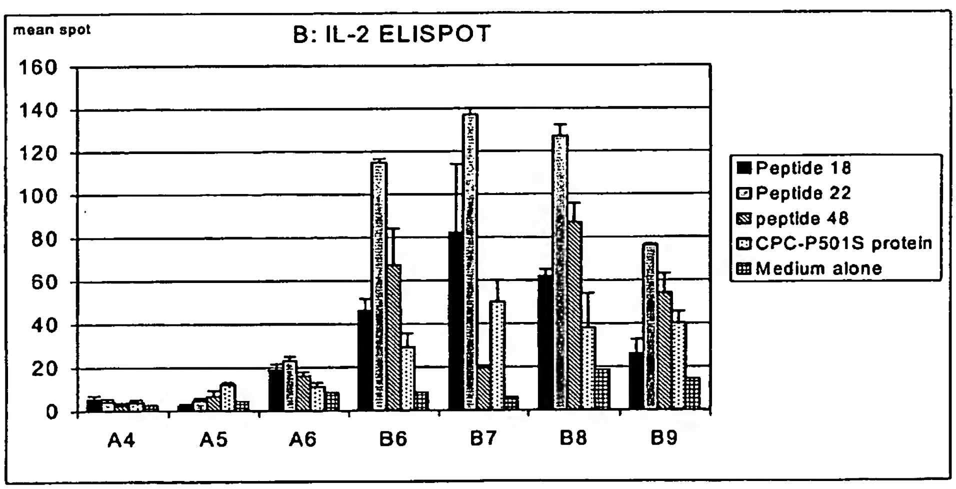
Rec'd PCT/PTO 08 DEC 2004

10/517420

FIG. 24 – Cellular responses by ELISPOT at day 77 following PMID immunisation at day 0, 21, 42, and 70 with pVAC-P501S (JNW680, B6-9) and pVAC empty (A4-6).

Graph A shows the IFN-γ responses whilst Graph B shows the IL-2 responses.





10/517420

FIG. 25 – Comparison of P501S and CPC-P501S.

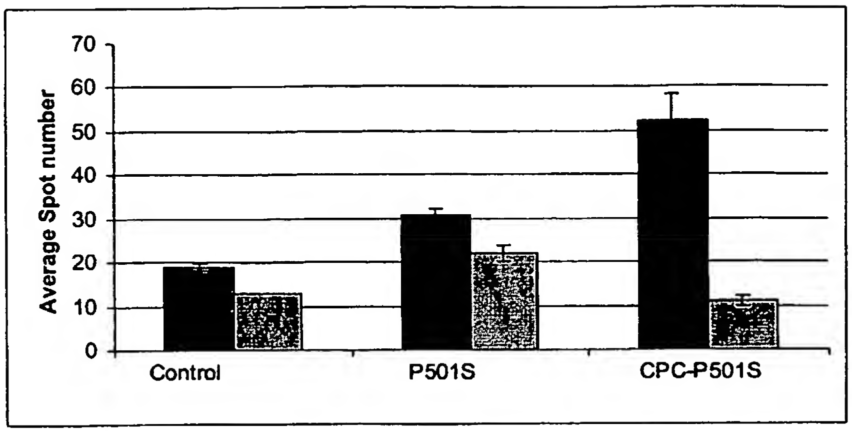
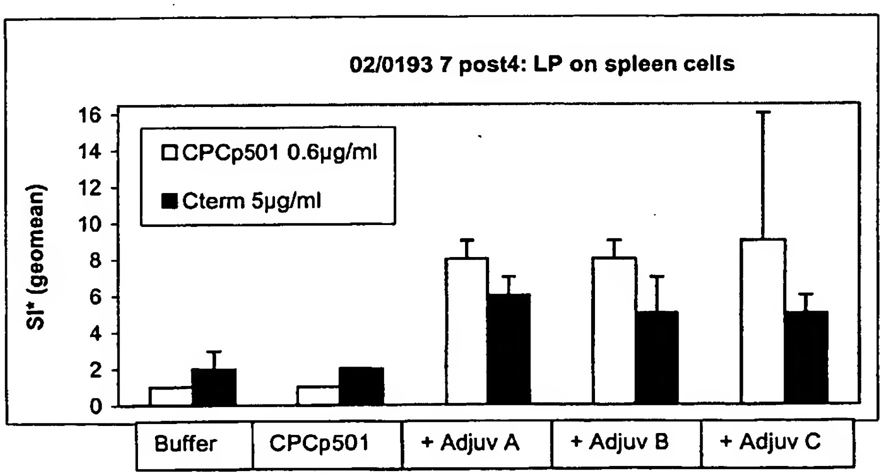


FIG. 26 – Immune response (lymphoproliferation on spleen cells) following protein immunisation with CPC-P501S.



Rec'd PCT/PTO 08 DEC 2004 10/517420

FIG. 27 – Evaluation of the immune response to different CPC-P501S constructs

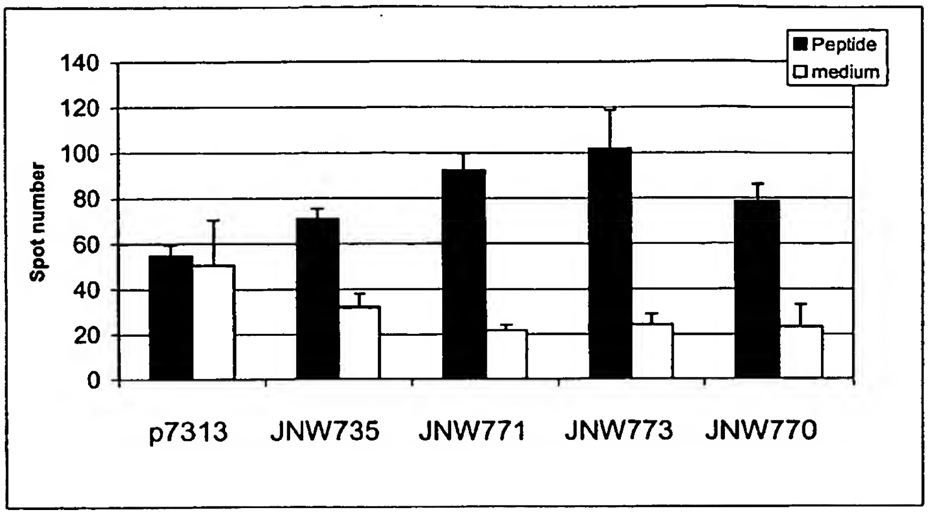


FIG.28. MUC1-CPC DNA and polypeptide sequences

FIG. 28A. DNA sequence (SEQ ID NO.49)

ATGACACCGGGCACCCAGTCTCCTTTCTTCCTGCTGCTGCTCCTCACAGTGCTTACAGTTGTTACAGGTTCTG
GTCATGCAAGCTCTACCCCAGGTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCCAGCTCTAC
TGAGAAGAATGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCGGTTCAGGCTCCTCCACCACT
CAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGGATG
TCACCTCGGTCCCAGTCACCAGGCCAGCCCTGGGCTCCACCACCCCGCCAGCCCACGATGTCACCTCAGCCCC
GGACAACAAGCCAGCCCCGGGCTCCACCGCCCCCCCAGCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCG
CCCCCGGGCTCCACCGCCCCCCCAGCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGCCCCCGGGCTCCA
CCGCGCCCGCAGCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCAGC
CCATGGTGTCACCTCGGCCCCGGACAACAGGCCCGCCTTGGCGTCCACCGCCCCTCCAGTCCACAATGTCACC
TCGGCCTCAGGCTCTGCATCAGGCTCAGCTTCTACTCTGGTGCACAACGGCACCTCTGCCAGGGCTACCACAA
CCCCAGCCAGCAAGAGCACTCCATTCTCAATTCCCAGCCACCACTCTGATACTCCTACCACCCTTGCCAGCCA
TAGCACCAAGACTGATGCCAGTAGCACTCACCATAGCACGGTACCTCCTCTCACCTCCTCCAATCACAGCACT
TCTCCCCAGTTGTCTACTGGGGTCTCTTTCTTTTTCCTGTCTTTTCACATTTCAAACCTCCAGTTTAATTCCT
CTCTGGAAGATCCCAGCACCGACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGCAGATTTA
TAAACAAGGGGGTTTTCTGGGCCTCTCCAATATTAAGTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCTG
GCCTTCCGAGAAGGTACCATCAATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAAACGGAAGCAGCCT
CTCGATATAACCTGACGATCTCAGACGTCAGCGTGAGTGATGTGCCATTTCCTTTCTCTGCCCAGTCTGGGGC
TGGGGTGCCAGGCTGGGGCATCGCGCTGCTGGTGCTGGTCTGTGTTCTGGTTGCGCTGGCCATTGTCTATCTC
ATTGCCTTGGCTGTCTGTCAGTGCCGCCGAAAGAACTACGGGCAGCTGGACATCTTTCCAGCCCGGGATACCT
ACCATCCTATGAGCGAGTACCCCACCTACCACACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTAG
CCCCTATGAGAAGGTTTCTGCAGGTAATGGTGGCAGCAGCCTCTCTTACACAAACCCAGCAGTGGCAGCCACT
TCTGCCAACTTGATGGCGGCCGCTTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCA
ATGGCACTTGGTACTACTTTGACAGTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAA
CTGGTACTGGTTCGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTC
AACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAG
GCGCCATGCAATACATCAAGGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTT
TATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAATGA

FIG. 28B. MUC1-CPC polypeptide sequence (SEQ ID NO.50)

MTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHSPGSGSSTT
QGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNKPAPGSTAPPAHGVTSAPDTRP
PPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTSAPDTRPAPGSTAPPAHGVTSAPDNRPALASTAPPVHNVT
SASGSASGSASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDASSTHHSTVPPLTSSNHST
SPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTL

AFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLVLVCVLVALAIVYL
IALAVCQCRRKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSTDRSPYEKVSAGNGGSSLSYTNPAVAAT
SANLMAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYYF
NEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQSADGTGWYYLKPDGTLADRPE

FIG.29. ss-CPC-MUC1 construct and sequence

FIG. 29A. DNA sequence (SEQ ID NO.51)

ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAGGTGTCCACTCCCAGGTCCAAATGGCGG
CCGCTTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTT
TGACAGTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAAC
TCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGA
AGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGCAATACATCAA
GGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGA
ACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAATGACACCGGGCACCCAGTCTC
CTTTCTTCCTGCTGCTGCTCCTCACAGTGCTTACAGTTGTTACAGGTTCTGGTCATGCAAGCTCTACCCCAGG
TGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCCAGCTCTACTGAGAAGAATGCTGTGAGTATG
ACCAGCAGCGTACTCTCCAGCCACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGG
CCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGGGGACAGGATGTCACCTCGGTCCCAGTCACCAG
GCCAGCCCTGGGCTCCACCACCCCGCCAGCCCACGATGTCACCTCAGCCCCGGACAACAAGCCAGCCCCGGGC
TCCACCGCCCCCCCAGCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGCCCCCGGGCTCCACCGCCCCCC
CAGCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGCCCCCGGGCTCCACCGCGCCCGCAGCCCACGGTGT
CACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCAGCCCATGGTGTCACCTCGGCCCCG
GACAACAGGCCCGCCTTGGCGTCCACCGCCCCTCCAGTCCACAATGTCACCTCGGCCTCAGGCTCTGCATCAG
GCTCAGCTTCTACTCTGGTGCACAACGGCACCTCTGCCAGGGCTACCACAACCCCAGCCAGCAAGAGCACTCC
ATTCTCAATTCCCAGCCACCACTCTGATACTCCTACCACCCTTGCCAGCCATAGCACCAAGACTGATGCCAGT
AGCACTCACCATAGCACGGTACCTCCTCTCACCTCCTCCAATCACAGCACTTCTCCCCAGTTGTCTACTGGGG
TCTCTTTCTTTTTCCTGTCTTTTCACATTTCAAACCTCCAGTTTAATTCCTCTCTGGAAGATCCCAGCACCGA
CTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGCAGATTTATAAACAAGGGGGTTTTCTGGGC
CTCTCCAATATTAAGTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCTGGCCTTCCGAGAAGGTACCATCA
ATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAAACGGAAGCAGCCTCTCGATATAACCTGACGATCTC
AGACGTCAGCGTGAGTGATGTGCCATTTCCTTTCTCTGCCCAGTCTGGGGCTGGGGTGCCAGGCTGGGGCATC
GCGCTGCTGGTGCTGGTCTGTGTTCTGGTTGCGCTGGCCATTGTCTATCTCATTGCCTTGGCTGTCTGTCAGT
GCCGCCGAAAGAACTACGGGCAGCTGGACATCTTTCCAGCCCGGGATACCTACCATCCTATGAGCGAGTACCC
CACCTACCACACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTAGCCCCTATGAGAAGGTTTCTGCA
GGTAATGGTGGCAGCAGCCTCTCTTACACAAACCCAGCAGTGGCAGCCACTTCTGCCAACTTGTAG

FIG. 29B. ss-CPC-MUC1 protein sequence Polypeptide sequence (SEQ ID NO.52)

MGWSCIILFLVATATGVHSQVQMAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDN
SGEMATGWKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQSADG
TGWYYLKPDGTLADRPEMTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSM
TSSVLSSHSPGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNKPAPG
STAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTSAPDTRPAPGSTAPPAHGVTSAP
DNRPALASTAPPVHNVTSASGSASGSASTLVHNGTSARATTTPASKSTPFSIPSHHSDTPTTLASHSTKTDAS
STHHSTVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLG
LSNIKFRPGSVVVQLTLAFREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGI
ALLVLVCVLVALAIVYLIALAVCQCRRKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSTDRSPYEKVSA
GNGGSSLSYTNPAVAATSANL